VOL. 12 No. 8, pp. 213-240 FEBRUARY 1951

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE OXIDATION OF HEAT-RESISTANT ALLOYS IN THE PRESENCE OF FOREIGN OXIDES

by J. L. MEIJERING and G. W. RATHENAU.

669.018.44:620.196:546.776-31

The construction of modern thermal engines with their high working temperatures has been made possible by the development of heat-resistant alloys possessing the mechanical properties required. Nevertheless cases still cocur where parts of these constructions show an unexpectedly rapid corrosion. From a detailed study of such phenomena of corrosion the authors have succeeded in throwing some light upon their causes.

In recent years more and more interest has been shown in heat-resistant alloys. These alloys have particularly become important for the construction of modern thermal engines (gas turbines, jet engines, hot air engines), in view of the fact that in order to reach a high efficiency the metallic parts of these engines must withstand working temperatures above 700 °C. This means that the materials to be used have to satisfy the requirement that, at the chosen working temperature and under the chosen mechanical load, during a reasonable lifetime for the engine particularly the plastic deformation and the corrosion due to chemical reaction with the surroundings do not exceed the given tolerances. This article will deal with a certain aspect of these phenomena, and in particular with the corrosion of heat-resistant metals and alloys through oxidation.

In principle there are two ways of avoiding oxidation of metallic parts of constructions at elevated temperatures.

In the first place "precious" metals can be chosen which, with respect to their surroundings, have too little affinity to oxidize. This means that at the working temperature the oxygen pressure (dissociation pressure) of the oxides that might be formed from the metal is greater than the partial pressure of oxygen in the atmosphere surrounding the metal. For instance, the oxygen pressure of silver oxide Ag_2O above 150 °C is more than $^1/_5$ atmosphere, so that above that temperature silver does not oxidize in air of atmospheric pressure.

However, owing to their unfavourable mechanical properties and high prices, precious metals (or alloys thereof) are hardly ever used for heat-resistant constructional parts. Most lythe second course is followed, basic or alloy elements being chosen which have the property of forming on the metal dense and strongly adhering oxide skins practically impervious to oxygen- and metal-ions. The alloy underneath such a skin certainly has the tendency to oxidize at the working temperature, but the rate at which this oxidation takes place is reduced to sufficiently small proportions by the obstruction to diffusion. We could express it in this way, that in the case of precious metals it is the thermodynamics (position of the dissociation equilibrium of the oxide) that make the material resistant to oxidation, whereas in the case of the protective oxide skin only the kinetics of the oxidation are affected in a sense favourable for the purpose.

As regards the protection afforded by an oxide skin, in many cases occurring in practice account has to be taken of the fact that at the surface the material comes into contact not only with oxygen or air but also with foreign substances, for instance oxides having an entirely different composition than that of the oxide skin making the material resistant to oxidation. It has long been known that such a contact sometimes considerably reduces this resistance to oxidation. An example of this is given in electric heating wires surrounded by certain ceramic materials. The smallest trace of asbestos in such a material may cause a wire of FeCrAlCo 73/20/6/1 [1]) ("Kanthal") heated to 1300 °C to fuse in a few seconds, whereas the normal lifetime of this wire at that temperature is some hundreds of hours. Another example is the abnormal corrosion of stainless steel in combustion chambers fired with fuel oil [2]). In this case it is the vanadium oxide so often present in such oils that is the cause of the trouble; this oxide may amount to 80% of what is left from the combustion of the oil. A third example relates to stainless steels containing molybdenum, from which, at the working temperature, volatile molybdenum oxide MoO_3 is formed [3]). Here again there is abnormal corrosion, which is attributable to the reaction of the oxide skin of the metal with the MoO_3 vapour.

Similar phenomena have been encountered by the authors when investigating in the Philips Laboratories at Eindhoven the oxidation-resisting properties of heat-resistant copper alloys and steels at 750 °C in air strongly contaminated with MoO_3. In the case of aluminium bronze (AlCu 8/92), for instance, heated in a furnace contaminated with MoO_3 vapour, after brushing off the oxide formed, which adhered loosely to the surface, a loss in weight was measured which was many hundreds of times greater than the loss found after oxidation in pure air. The graph in *fig. 1* shows that after three hours in the contaminated atmosphere the loss in weight already amounted to 6 mg/cm², as compared with about 0.01 mg/cm² in pure air. A remarkable fact shown by this graph is that the corrosion continues at a rapid rate when the material is further heated in a pure atmosphere (thus not containing MoO_3). We shall revert to this detail later (see fig. 5*b*).

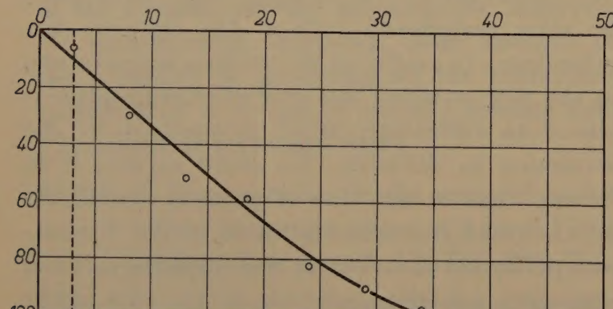

Fig. 1. Oxidation of a strip of aluminium bronze, AlCu 8/92 (cf. footnote [1])), heated at 750 °C in air contaminated with vapour from molybdenum oxide MoO_3. The loss of weight, measured after brushing off the loose oxide, is plotted as a function of time. — After three hours' heating the strip was transferred into an atmosphere of pure air; the oxidation appeared to continue at about the same rate.

With a view to gaining a general insight into the manner in which the oxidation of metals is influenced by contact of the protective oxide skin with foreign substances, in particular oxides, a series of systematic experiments have been carried out in this laboratory. In all cases MoO_3 was used as foreign oxide, but, as will appear from what follows, the results and the conclusions to be drawn therefrom may be regarded as more or less representative for the mechanism of corrosion also when other foreign oxides are active, and probably also in the case where a metal is corroded not by air but by other gases, as for instance chlorine.

Quantitative determination of the degree of oxidation

First of all the process of oxidation was investigated, by different means, as a function of temperature. It very soon became evident that abnormal oxidation of heat-resistant alloys caused by MoO_3 may occur already at relatively low temperatures, even below 500 °C. This was established by bringing samples of the metals and alloys into intimate contact with MoO_3 powder during the heating. Contact only with air containing MoO_3 does not, at these temperatures, cause any appreciable rapid corrosion, since the vapour pressure of MoO_3 at 500 °C is only about 10^{-5} mm Hg. For a proper mutual comparison, in all the tests the samples were therefore packed in MoO_3 powder and then heated in air. Reference samples of the same composition, carefully separated from the others, were heated to the same temperature in pure air alone.

In order to follow quantitatively the conversion through oxidation, in the test to which fig. 1 refers the loss in weight of the metal was taken as the measure. It appears, however, that in some cases this leads to difficulties, because sometimes it is difficult to remove the oxide skin entirely and, furthermore, the abnormal corrosion is often only local (as will be understood from what follows). Both these factors cause the results to be spread over a wide range, as is clearly seen from *fig. 2*, where the loss in weight measured with a series of strips of the same aluminium bronze as used for the test of fig. 1 — each strip heated for 3 hours in contact with MoO_3 and air — has been plotted as a function of the temperature. It is seen that round about 480 °C oxidation strats to take place much more quickly, but from results with such a considerable spread it is difficult to draw more detailed conclusions as to the relation between the rate of oxidation and temperature.

---

[1]) An alloy with 73 wt % Fe, 20 wt % Cr, 6 wt % Al, 1 wt % Co. Further in this article the abbreviated form will be used in every case.

[2]) P. Schläpfer, P. Amgwerd and H. Preis, Schweizer Archiv **15**, 291, 1949.

[3]) W. C. Leslie and M. G. Fontana, Trans. Am. Soc. Metals **41**, 1213, 1949.

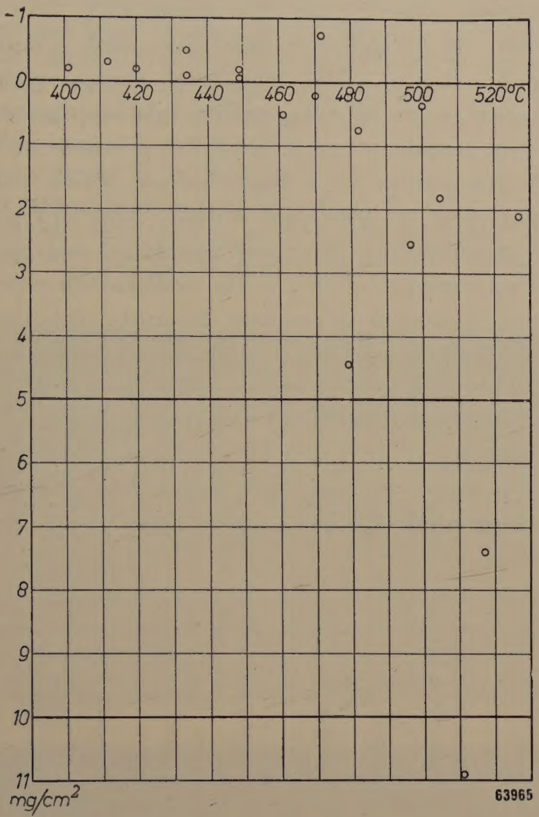

Fig. 2. Loss in weight of strips of AlCu 8/92 packed in MoO_3 powder and heated for three hours in air at different temperatures.

A more reliable and more practicable method for determining the degree of oxidation was found to be the following.

From each material to be investigated a set of five wires were used (diameter $2r_0 = \frac{1}{2}$ mm) and brought to a reproducible, stress-free, initial state by heating in a pure mixture of nitrogen and hydrogen. After oxidation in contact with MoO_3 and air, the average breaking strength (K_M) of the wires was determined at room temperature with a normal tensile testing machine. For comparison an identical set of wires were heated in pure air for the same length of time and at the same temperature, after which their average breaking strength (K_A) was likewise measured. As a measure for the abnormal oxidation we have now defined an "effective penetration depth" d:

$$d = (1 - \sqrt{K_M/K_A})\, r_0 \; . \quad . \; . \; . \; . \qquad (1)$$

The definition is based upon the assumption that the wires subjected to the tensile test always break at the same tensile stress and that no appreciable notch effects occur. This means to say that all wires break at one and the same value of the quotient of breaking strength and area of the remaining cross section of the wire. If the oxidation round the circumference of the wire penetrates to a depth d then the radius r_0 of the effective wire cross section is reduced to $r_0 - d$, from which the above formula follows directly (provided, as is indeed found to be the case, corrosion in pure air is negligible).

The penetration depth d relates to the most corroded spots, since a test specimen usually breaks at the thinnest place. Therefore, although in some places the oxidation may be impeded through incidental causes (see below), with this method of measuring and this definition reproducible results may still be expected. Incidentally, it may be noted that the penetration depth measured by the breaking strength is quite a good measure for corrosion in practice, since in practical engineering it is just the deterioration in mechanical strength of constructional parts, be it only local, that counts most. For a good understanding of the mechanism of corrosion, however, it is certainly also of importance to know whether and in how far there is question either of all-round corrosion or of a local, e.g. inter-crystalline, penetration. For this reason also cross sections of the wires were examined microscopically.

The fact that really reproducible values are found for the penetration depth is demonstrated by *fig. 3*, showing the penetration depth d determined according to eq. (1) after 3 hours' heating and plotted as a function of temperature, for the same aluminium bronze AlCu 8/92 to which figures 1 and 2 relate. Here the measured points show but little spread and the curve determined by the points clearly shows two discontinuities, the first at 470 °C and the second round about 535 °C.

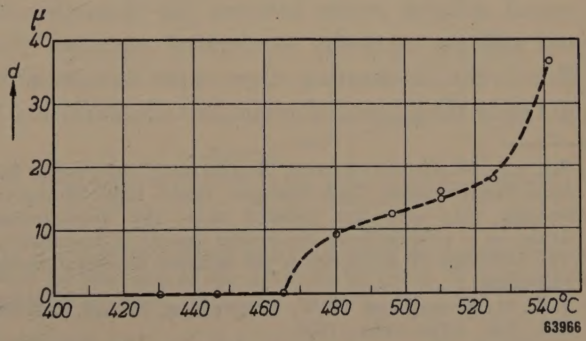

Fig. 3. Penetration depth d found from tensile tests with wires of AlCu 8/92 oxidized (in the same way as the strips of fig. 2) by heating for three hours in MoO_3 powder and air, at different temperatures.

Liquefaction at and in the protective oxide skin

The fact that at very definite temperatures the rate of oxidation changes discontinuously gives rise to the thought that phase transitions, namely melting or a change in the crystal structure, will play a part. In itself this is no new idea. The possibility of liquefaction playing a part in accelerated oxidation has been suggested before [4]. The examples given in the beginning also point in that direction: asbestos, which may cause rapid corrosion of heating wires, melts at about the temperature of 1300 °C mentioned; vanadium oxide, which occurs in fuel oil, melts already at about 650 °C. The melting point of molybdenum oxide lies, it is true, much higher than the temperature of about 500 °C at which abnormal corrosion was found, namely at 795 °C; but even so liquefaction is possible at those lower temperatures if low-melting eutectics occur.

To put this idea to the test we have determined the eutectic temperatures of the oxide systems concerned, with the aid of temperature-time curves (when a melt is cooled a retardation in the decrease of the temperature points to a separation of crystals, and if the cooling comes to a standstill for a time this points to a eutectic temperature having been reached, at which the whole of the liquid solidifies, as can in fact be seen to take place). In the case of the aforementioned aluminium bronze in combination with MoO_3 we found that the binary system [5] MoO_3-Cu_2O has a eutectic temperature of 535 °C. The eutectic temperature of the ternary system MoO_3-MoO_2-Cu_2O proved to be 470 °C (that of MoO_3-MoO_2 lies at 778 °C).

Not only do we find here, therefore, very low melting eutectics, but the striking agreement between the eutectic temperatures — thus the lowest temperatures at which liquids can occur at the surface of the alloy — and the discontinuities in fig. 3 clearly lends support to the hypothesis that a causal relation exists between the formation of liquid and the markedly accelerated corrosion [6][7].

How is the accelerating effect of the formation of liquid upon the process of oxidation to be explained? It has to be imagined that the liquid, even if present in only small quantities, forms continuous channels along the grain boundaries of the protective oxide, and that owing to diffusion and convection in these liquid channels, which permit a much more rapid exchange of metal atoms and especially also of oxygen atoms than is possible in a solid phase, the protection afforded by the oxide skin becomes illusory. It is quite feasible that such channels may be formed. The interfacial energy of two adjacent grains of a crystalline phase (the oxide skin) is usually so great that the presence at the grain boundaries of a liquid phase which easily adapts itself to the local atomic structure leads to a reduction of the total interfacial energy. A striking example of this is the well-known case of copper immersed in molten bismuth, the liquid penetrating along the surfaces of the adjacent grains deep into the metal and thereby making it extremely brittle [8]. In many other cases, where the interfacial energies at the grain boundaries may perhaps be less, the liquid may anyhow penetrate along the ribs where three grains come together.

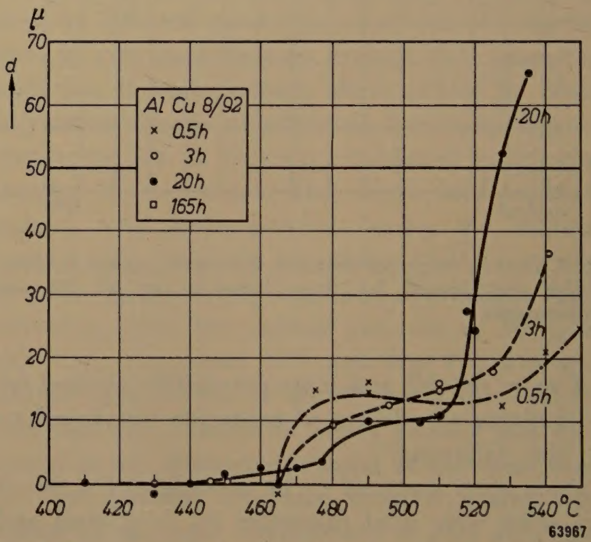

Fig. 4. The same as in fig. 3 but also after ½hour, 20 hours and 165 hours' heating respectively. At not too high temperatures there is a peculiar dependency of the penetration depth upon the heating time. This is accounted for by a notch effect at the beginning of the oxidation.

Details of the formation of the eutectics

In the case dealt with in the foregoing, it is peculiar that the low-melting eutectic is not formed by interaction of the added oxide MoO_3 with the "protecting" oxide Al_2O_3 but by reaction with oxides of the basic metal, copper. (The oxidation of pure aluminium is not accelerated by MoO_3!) The possibility of such a reaction is probably to be ascribed to the fact that in the oxide skin of the AlCu oxidized in air there are also Cu-oxides present, whereas the MoO_3 could also react directly with the Cu from the alloy underneath through cracks in the protective layer of oxide. If such capillary cracks do indeed play a part then it is understandable why the reaction is often localized, as mentioned above in connection with the definition of the penetration depth [9]). This may also be the explanation of the fact that, curiously enough, the penetration depth at first appears to decrease with increasing duration of heating. *Fig. 4* illustrates this effect. If, owing to its starting from fine cracks in the oxide skin, the corrosion at the beginning is greatly limited locally, then it lowers the breaking strength disproportionately owing to notch effect. After a greater length of time the corrosion becomes more uniform, as *figs 5a* and *b* show.

---

[4] See, e.g., W. Hessenbruch, Metalle und Legierungen für hohe Temperaturen, I., J. Springer, Berlin 1940. Strangely enough this idea has receded into the background again, as is evident from the article quoted in footnote [3].

[5] The influence of Al_2O_3 upon the eutectic temperatures is negligible.

[6] J. L. Meijering and G. W. Rathenau, Nature, London **165**, 240, 11th Febr. 1950.

[7] G. W. Rathenau and J. L. Meijering, Metallurgia, **42**, 167-172, 1950. This article gives a more detailed report on the investigations dealt with here, in particular also on tests with various alloys not mentioned here and on the results of the metallographic examination.

[8] C. S. Smith, Trans. A. I. M. E. **175**, 15, 1948.

The influence of MoO_3 upon the oxidation of pure Cu is to be deduced from *fig. 6*. In this case it is of course one of the protecting oxides itself that reacts with the foreign oxide, namely Cu_2O (the eutectic temperature for CuO with MoO_3 is 735 °C). The influence of the eutectic point at 535 °C of MoO_3-Cu_2O is quite clear. Here the ternary eutectic point at 470 °C of MoO_3-MoO_2-Cu_2O is not manifest, though its influence is to be seen in measurements of the loss of weight, which are more sensitive than those of the penetration depth, when the object is to detect the very beginning of accelerated oxidation. The fact that here the formation of MoO_2 from MoO_3 is not noticeable, whereas it is so in the case of Al-Cu alloys, is to be accounted for by the strongly reducing action of Al (great stability of Al_2O_3).

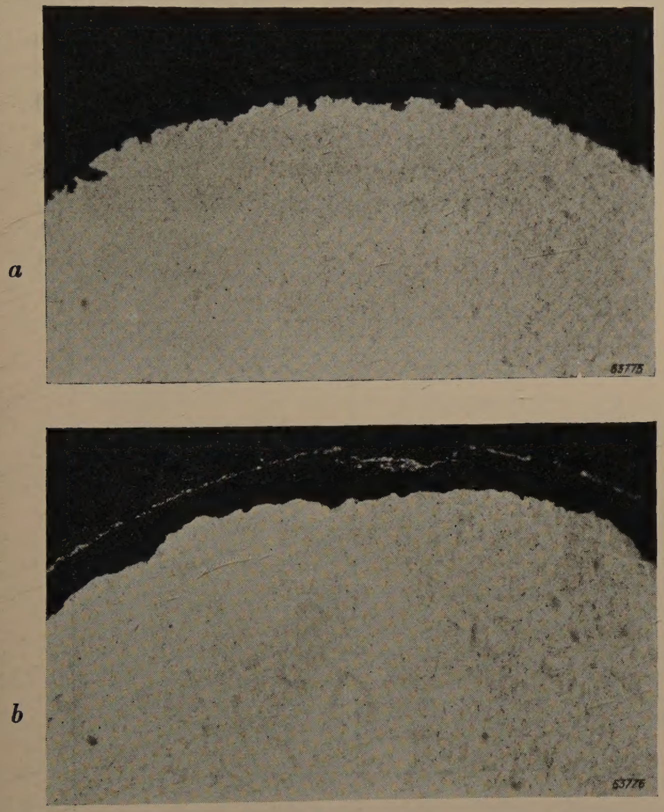

Fig. 5. *a*) Microphotograph of a specimen of the weakly etched cross section of a wire of AlCu 8/92 after heating for half an hour at 510 °C in MoO_3 and air. The black surrounding is synthetic resin ("Philite") which was pressed round the wire to facilitate flat grinding of the surface of the cross section. The corrosion on the surface of the wire is strongly localized (notch effect in the tensile test). Enlargement 600 times.
b) The same as *a*) but after three hours' heating of the wire. The corrosion here is decidedly more uniform. In the case of this specimen there is a light ring round the outside of the wire. In this layer, as revealed by X-ray diffraction examination, a layer of pure Cu has been formed which apparently has been precipitated from the liquid oxide mixture during the cooling of the wire after heating. Underneath this layer is a darker ring consisting mainly of molybdenum oxides (again found from X-ray diffraction). These oxides are responsible for the further rapid oxidation of the bare wire transferred, after removal of the loose oxide skin, to an atmosphere of pure air and further heated (cf. fig. 1).

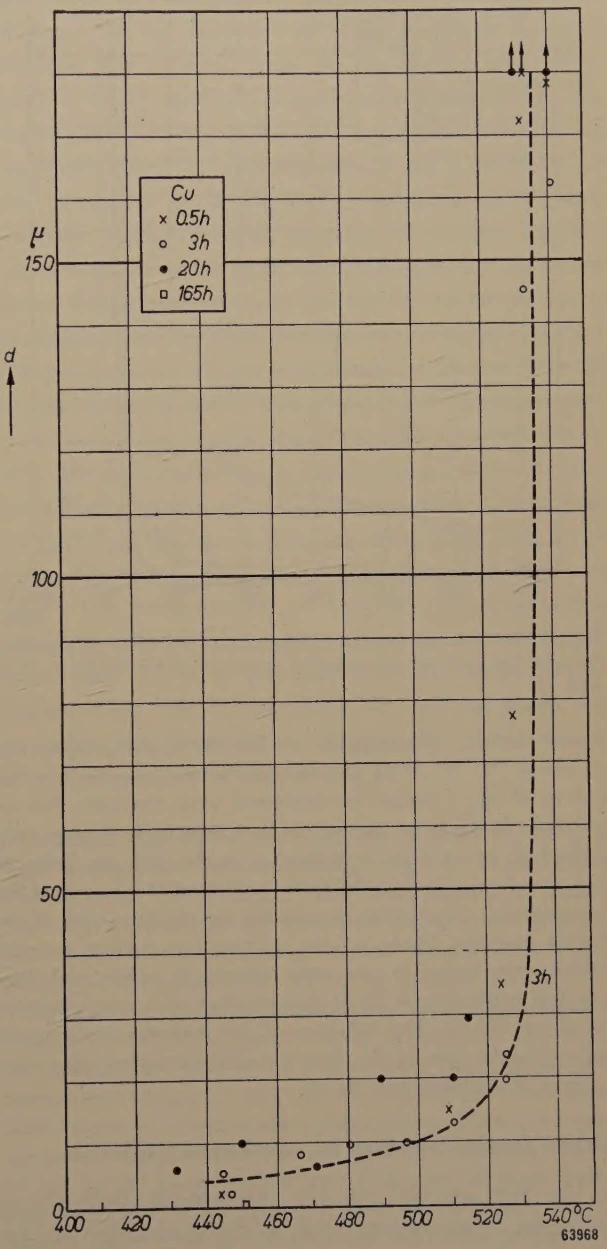

Fig. 6. Penetration depth as a function of the heating temperature in the case of the corrosion of wires of pure copper (OFHC) in MoO_3 and air.

---

[9]) Owing to droplets flowing together and forming larger drops the supply of MoO_3 may be locally interrupted, the corrosion then being localised to such an extent that it becomes very irregular.

Also with heat-resistant alloys of **iron** the simple case occurs of considerably increased oxidation taking place at the temperature at which the protective oxide itself reacts with the added oxide under the formation of liquid. In *fig. 7* the penetration depths have been plotted for chromium-iron CrFe 25/75. The greatly increased oxidation above 770 °C corresponds to the liquefaction of the eutectic MoO_3-MoO_2-Cr_2O_3. Of all the possible eutectics in the whole system this is not the lowest melting one: according to our measurements the lowest eutectic temperature, of FeO-MoO_3-MoO_2-Cr_2O_3, lies at about 705 °C. Why this eutectic is **not** manifest in the curves of fig. 7 cannot be explained with certainty, but in principle it might be due to the fact that there is only very little FeO present in the protecting oxide skin, too little, in the case of reaction with MoO_3-MoO_2, to reach the domain of the (quaternary) phase diagram suitable for the formation of the lowest eutectic. There is also another low-melting eutectic which may occur in the said system of oxides and does not find expression in fig. 7, namely that with a temperature of about 725 °C. The influence of this eutectic, however, is quite clearly seen in the case of a chromium-iron alloy containing 9% nickel; see *fig. 8.*

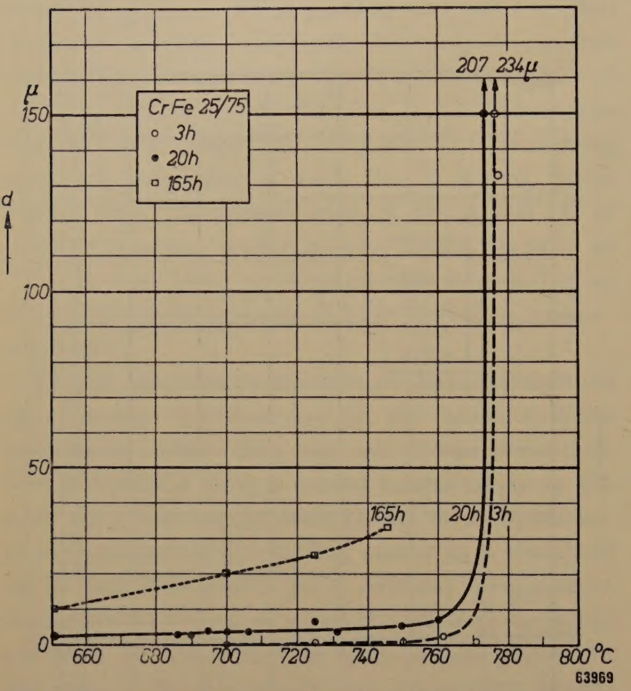

Fig. 7. Penetration depth as a function of the heating temperature in the case of corrosion of wires of CrFe 25/75 in MoO and air.

Other factors tending to accelerate oxidation

On the ground of the agreement between a eutectic temperature and a discontinuity in the relation of penetration depth and temperature, the formation of liquid in the oxide skin is to be regarded as the cause of the accelerated oxidation observed. The formation of liquid, however, cannot be the only cause. Figures 7 and 8, for instance, clearly show that the presence of MoO_3 also causes a more or less gradual increase of oxidation of chromium-iron alloys below 705 °C, which in this case, as already mentioned, was the lowest eutectic temperature.

Now an acceleration of oxidation is to be expected also in the absence of liquid if molybdenum ions from the molybdenum oxide were taken up in the lattice of the protective oxide skin. It may be imagined that a lattice containing Mo-ions will not be stoichiometric, since Mo may quite easily occur in different valencies side by side. This means to say that in the lattice there will be unoccupied places making a relatively easy transportation of metal- or oxygen-ions possible, whilst on the other hand owing to an interchange of molybdenum-ions of different valencies a high order of mobility of electrons is ensured. Since, according to the theoretical views, obtained particularly by the work done by Wagner [10]) and by Mott [11]), in the oxidation of metals and alloys it is to be presumed that there is a separate diffusion of metal-ions and electrons through the oxide skin, in that skin with non-stoichiometric lattice we see literally a path free for the rapid oxidation of the underlying metal [12]).

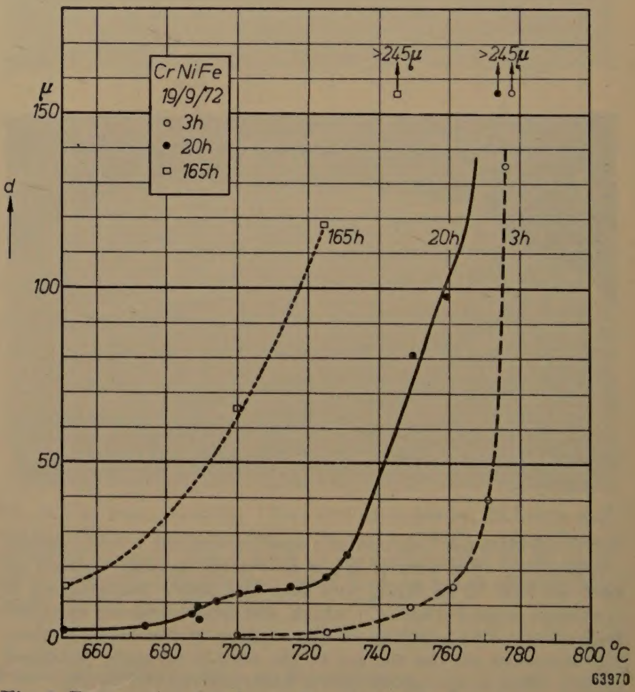

Fig. 8. Penetration depth as a function of the heating temperature in the case of corrosion of wires of CrNiFe 19/9/72 (stainless chromium-nickel steel) in MoO_3 and air.

---

[10]) C. Wagner, Z. phys. Chem. B **21**, 25, 1933; **32**, 447, 1936.

[11]) N. F. Mott, in N. Cabrera and N. F. Mott, Rep. on Progr. in Phys. **12**, 163, 1949.

[12]) Attention is drawn, in passing, to the parallelism between good protective action and poor electric conductivity of the oxide skin; for the latter point see Verwey, Haayman and Romeyn, Philips Techn. Rev. **9**, 239, 1947/1948.

Absorption of a small percentage of molybdenum-ions in the lattice of the very thin oxide skin would be very difficult to prove experimentally. In the case considered here, however, the change brought about by the presence of MoO_3 appeared to be far more drastic: X-ray diffractional examination of the material in the oxide skin of the chromium-iron alloy CrFe 25/75 showed that after 165 hours' heating in MoO_3 and air at 650 °C there was no longer any sign of the structures normally to be expected (Cr_2O_3 and $FeO \cdot Cr_2O_3$), nor of Fe_3O_4, FeO and suchlike. Exclusively a new structure was found, the lattice type of which has not yet been ascertained but which, judging from a spectroscopic analysis, belongs to an oxide consisting for the greater part of molybdenum oxides (12% FeO or Fe_2O_3, 6% Cr_2O_3 and the rest Mo-oxides). Owing to its relatively high permeability, what there is of oxide skin cannot, therefore, afford any effective protection to the metal.

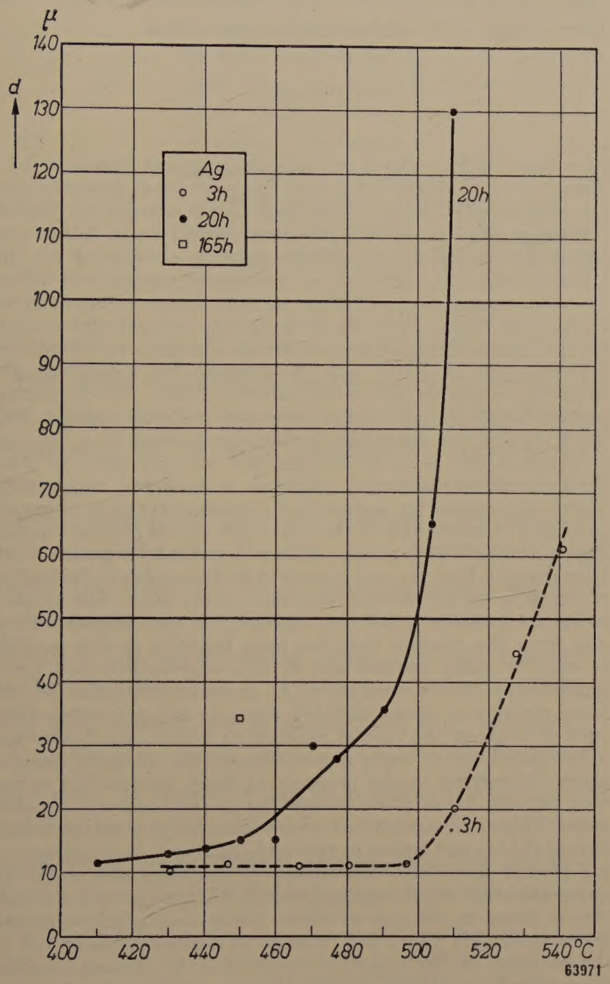

Fig. 9. Penetration depth as a function of the heating temperature in the case of corrosion of 99.6% pure Ag in MoO_3 and air. The silver wires were first preheated for 12 minutes at 600 °C in a dry mixture of nitrogen and hydrogen entirely free of oxygen.

Yet another mechanism has to be assumed to account for our findings in the case of oxidation of silver and silver alloys. Experiments with samples of silver packed in MoO_3 powder showed at about 300 °C a noticeable depth of penetration increasing gradually up to 490 °C and then more pronounced at still higher temperatures; see *fig. 9*. At first sight the corrosion is astonishing, because, as mentioned in the introduction, above 150 °C in air of normal pressure silver cannot oxidize at all on thermodynamic grounds, owing to the great dissociation pressure of Ag_2O. In the presence of MoO_3 however double oxides are formed, such as Ag_2MoO_4, and these have a much smaller dissociation pressure [13]). We found that, even at its melting point of 570 °C, Ag_2MoO_4 does not decompose in air. Therefore, in the presence of MoO_3 silver behaves as a non-precious metal, which is only protected against rapid oxidation at high temperatures by an oxide skin. Therefore, we find in fig. 9, as in previous examples, the marked increase of oxidation at that temperature at which a eutectic in the oxide skin becomes liquid (495 °C is the lowest eutectic temperature in the system Ag_2MoO_4-MoO_3), and finally also strong oxidation in the case of contact only with MoO_3 vapour at temperatures where the vapour pressure of MoO_3 is not too low (750 °C). It is seen, therefore, that by the addition of MoO_3 to Ag_2O the temperature at which the dissociation pressure exceeds 1/5 atmosphere is raised by at least 600 °C. It is possible that at temperatures higher than 750 °C, to which we did not go in our experiments, there is no longer any oxidation of silver in air containing MoO_3.

Although this makes the oxidation of silver by MoO_3 at high temperatures understandable, there is a remarkable feature about this oxidation that needs explaining, namely the small temperature-dependency of the measured penetration depth below the lowest eutectic temperature. When, instead of the penetration depth, we measure the loss in weight as function of the temperature, then the temperature-dependency below the lowest eutectic temperature appears to be much greater. This would indicate that the "temperature-insensitive" corrosion below 490 °C must, at least for a large part, be due to intercrystalline penetrations, which at increasing temperature do not increase so much in depth as in number or in width. Such penetrations have indeed been found microscopically in specimens of wires heated below 490 °C. (Above 490 °C the penetrations are still more pronounced; see *fig. 10*.)

Such a penetration, at temperatures where no liquid occurs and where the vapour pressure of MoO_3 is entirely negligible, is difficult to understand. It may be that owing to the reaction of Ag and MoO_3 in the solid state, and the attendant changes in volume, strong mechanical stresses arise in the silver crystals and cause fractures along the grain boundaries. Such a corrosive mechanism would have to be rather independent of temperature, but it would be strongly dependent upon small contaminations at the grain boundaries. Probably there is a relation between this and the fact that in the case of silver this corrosion appeared to be the more intense the less the degree of purity of the silver.

---

[13]) For analogous double oxides, e.g. of silver with vanadium or with chromium, something similar has been known for a long time: see R. Schenck, Z. anorg. allg. Chem. **249**, 88, 1942.

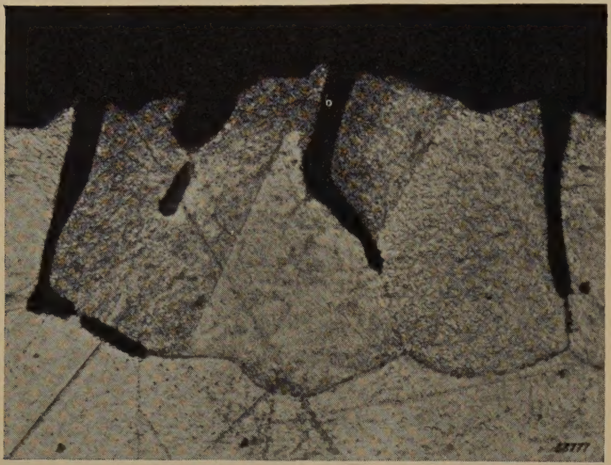

Fig. 10. Microphotograph of a specimen of the cross section of a block of silver (about 99.5% Ag) oxidized by heating for 20 hours at 530 °C in MoO_3 and air. There is strong intercrystalline corrosion, at least along the boundaries of grains orientated at random. On the straight boundaries separating twin parts of a grain the crystalline structures to the right and left are exactly matched. The interfacial energy on these boundaries is thus very small, so that the liquid (formed at 495 °C) does not penetrate there. Also these twin boundaries have been made clearly visible by etching the specimen with $NH_4OH + H_2O_2$. Enlargement 500 times.

The three effects discussed here: formation of a liquid, formation of a non-stoichiometric oxide lattice, and displacement of a thermodynamic dissociation equilibrium, probably play a part in many hitherto unaccountable cases of corrosion of heat-resistant alloys in the presence of foreign substances. Though there may, of course, be still other important mechanisms at work, in all the cases of oxidation of alloys investigated by the authors the three mechanisms mentioned were sufficient to account satisfactorily, on broad lines, for the phenomena observed.

Summary. Alloys which are protected against corrosion by a dense skin of oxide, and thus may serve as heat-resistant constructional materials, sometimes show a strongly accelerated oxidation when brought into contact with foreign substances, especially foreign oxides. With a view to accounting for this phenomenon, the oxidation of a series of metals and alloys, i.a. with copper, iron and silver as base, has been investigated for the case where these are brought into contact with molybdenum oxide (MoO_3 in powder form). The degree of oxidation ("penetration depth") was determined by measuring the breaking strength of wires drawn from the respective material and heated in air with and without contact with MoO_3. When the oxidation is thus determined as a function of temperature a curve is obtained showing one or more discontinuities where the oxidation is suddenly accelerated. These discontinuities appear to lie at the eutectic temperatures of the systems of the respective oxides (lowest temperatures at which the formation of liquid occurs). From this it can be concluded that the accelerated oxidation is due to formation of liquid in the protective oxide skin of the alloy. The accelerating effect of the liquid upon oxidation is readily understood, and from the relation that has been found it is also possible to explain many peculiarities in the temperature- and time-dependency of the oxidation. In some cases oxides of the basic metal not essentially belonging to the protecting skin, such as Cu_2O in the case of aluminium bronze, may considerably contribute towards a lowering of the eutectic temperature. Sometimes in the presence of MoO_3 perceptible oxidation already takes place below the lowest eutectic temperature. This is accounted for by the absorption of molybdenum ions in the crystal lattice of the oxide skin and the accompanying greater freedom of movement of electrons and metal- or oxygen-ions. A third way in which MoO_3 may promote oxidation is found in the case of silver. Silver, as a precious metal, is not oxidizable above 150 °C in air of normal pressure, but in the presence of MoO_3 double oxides may be formed, making it possible for oxidation to take place at much higher temperatures; then again melting phenomena in the oxide skin make their influence felt.

ON THE REGULATION OF INDUSTRIAL PROCESSES

by H. J. ROOSDORP.

621-52

Regulating systems for industrial processes may be of a widely different nature: think, for instance, of the governor used on the oldest steam engines for keeping their speed constant, the voltage and frequency stabilizers for generators, the temperature regulators that are so highly important in the chemical industry, for distilling and cracking in the oil industry and for the preserving of foodstuffs. All these apparently quite different devices have one principle in common, that whatever it is that is to be kept constant is fed back via a controller upon a controlling quantity governing the first mentioned quantity.

It is the intention to follow up this article, giving a general review of such regulating systems, by others dealing in turn with a kind of electronic calculating machine for solving regulating problems and with a developed regulating device.

For most industrial processes it is necessary that some physical quantity, such as temperature, pressure, rate of flow, acidity or conductivity, should be kept at a certain level or value within a given tolerance. For this purpose there must be some regulating unit upon the position of which the quantity to be kept constant depends, as for example the gas valve of a gas-fired furnace in which the temperature must be kept constant. In the simplest cases it suffices to set the regulating unit in a certain position derived from theoretical calculations or practical experience (*fig. 1*). Allowance may be made as far as possible for the variations to be expected in all factors influencing the quantity to be regulated. In the case, for instance, of the furnace just mentioned, such factors are the charge in the furnace, the pressure of the gas, the ambient temperature, etc. During the "process" — in this case the generation of heat inside the furnace — there is no feedback of the temperature in the furnace upon the position of the gas valve. There is no provision for correcting, while the "process" is proceeding, any unexpected changes subsequently taking place in the working conditions and affecting the temperature. Regulating — if it may still be called that — is therefore limited, in such an open circuit, to choosing a predetermined position of the regulating unit as well as possible.

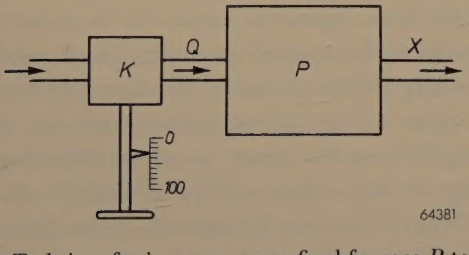

Fig. 1. To bring, for instance, a gas-fired furnace P to a certain temperature, with a valve K the gas flow Q can be pre-adjusted once for all so that the temperature X assumes and maintains as well as possible the desired value. Here there is no influence of the output quantity X upon the input quantity Q (open circuit).

A closed circuit (*fig. 2*) has a different behaviour. Here the quantity to be regulated is continuously measured and from information thus obtained the adjustment of the regulating unit to reach or maintain the conditions desired is deduced. Henceforth regulation is to be understood as referring only to this system. In this case there is a continuous influence (feedback) — making the circuit a closed one — of the output of the process upon the input. This feedback may be an observer (manual regulating system, fig. 2*a*) or an automatic controller (automatic regulating system, fig. 2*b*).

In what follows we shall only consider closed circuits with automatic controllers.

Accuracy and stability

The ideal regulating system would be that which works in such a way that the quantity to be regulated can never deviate from the desired value in spite of interfering, mostly random, variations in all sorts of working conditions which, without regulating system, would have a prohibitive effect upon the process. When it is desired to change over to some other value of the quantity to be regulated some adjustment has to be made to the controlling unit, and an ideal regulating system must also immediately respond to the adjustment.

It is due to the limited sensitivity and the inertia of various elements of the circuit that an actual regulating device does not fully answer the ideal in either of these two respects. By making the controlling unit highly sensitive it is of course possible to arrange that the slightest deviation from the desired value causes a considerable change in the position of the regulating unit, but owing to the time delay of the detecting element, of the controller and of the regulating unit, the latter does not reach its new position until some time after the disturbance has taken place, and in consequence of the (often very considerable) inertia of the process the effect of the said adjustment is delayed.

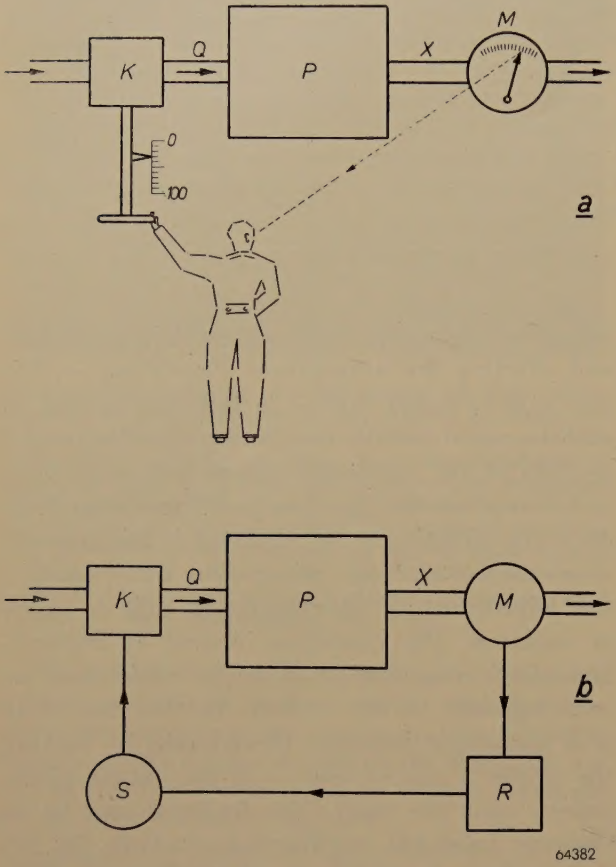

Fig. 2. *Closed* regulating circuits. The output quantity X to be regulated is measured by a detecting and measuring element M. According to the magnitude of the deviations from the desired value of X, the input quantity Q is adjusted, by means of the regulating unit K, through the intermediary of either (a) an observer or (b) an automatic controller R and a servomotor S.

After a disturbance — at least temporarily — there is consequently a certain difference between the actual and the desired value of the quantity to be regulated. When a new stationary state has been reached this deviation may in some cases be zero or the difference between the two values mentioned may be reduced to within certain limits depending, inter alia, upon the sensitivity of the detecting element (which may usually be ignored), but there are other cases where the deviation may remain at a certain value, and one then speaks of *static inaccuracy*, or *offset*.

The deviation which, as already said, is always present prior to the stationary state having been reached is called the *transient phenomenon*. Here there are three cases to be distinguished: (1) the *aperiodic* case, where the deviation always has the same sign while gradually decreasing, (2) the *damped periodic* case, where the deviation diminishes as the decay of a damped oscillation, and (3) the *non-damped periodic* case, when the deviation continues to fluctuate between certain limits without decaying.

The importance of offset is immediately obvious: it denotes in how far the quantity to be regulated approaches (in course of time) the desired value when a disturbance has taken place. But also the dynamic behaviour of the system is important: the deviations from the desired value following upon a disturbance might, for instance, lead to an intolerably large amplitude, or it might take far too long before a new stationary state is reached. It may even happen that the desired value is not reached at all owing to the input quantity undergoing every time another disturbance before any stationary state can be attained, each of these new disturbances affecting the output to be regulated continuously. As an example may be mentioned the regulating of the pressure in a drum filled with gas supplied by a piston pump. In such a case there is a strong pulsatory variation in the gas pressure following the frequency of the pump. This fluctuation, the amplitude of which may possibly reach intolerable proportions, is called the *dynamic inaccuracy*.

Aperiodic and damped periodic cases are said to be *stable* (since a stationary state is reached by the time a new disturbance arrives), while non-damped periodic cases — also referred to as *hunting* — are *unstable*. Farther on we shall see that a certain group of controllers (the so-called discontinuously acting controllers) are unstable by their very nature, so that there is always inevitably a certain fluctuation in the quantity to be regulated. If such a fluctuation is to be regarded as prohibitive a so-called continuously acting controller has to be used, which may under certain conditions be stable.

The two kinds of inaccuracy, the stability (in the case of continuously acting controllers) and the transient are the main factors to which particular attention must be paid in the regulating technique. They depend, of course, upon the properties of the elements comprising the closed circuit. Usually the plant is already defined and the regulating system has to be adapted to it in such a way that the desired accuracy and stability are obtained. If the requirements were to be worked out entirely by calculation then in many cases insurmountable difficulties would be encountered. On the other hand it is not advisable either to try to find the best regulating system solely by empirical means during the manufacturing process, because this takes much time and might upset the production for long periods. What has proved to be the best method in many cases is to study the situation with the aid of an electrical model.

Working with an electrical model

When working with a model the measurements to be taken from the plant itself can be limited to the recording of a step-function response curve representing the variations of the output quantity (which is to be regulated) as a function of time when the input is changed according to a certain function of time, for instance as a sudden change. Then an electrical network must be designed having a response curve, on a different time scale, corresponding as closely as possible to that of the process. This network then serves as a model, the behaviour of which, in combination with a model of the automatic controller, can be studied quickly and without interfering with the actual process. Guided by some experience one chooses for the model of the automatic controller one of the various types that are possible.

As a rule there is a considerable difference between the time scale of the actual process and that of the electrical network: a change requiring several minutes, for instance, in the process may take only a few milliseconds in the network. A periodic repetition of the sudden change makes it possible to observe the response curve of the network, and also the behaviour of the model with the automatic controller on the screen of an oscilloscope. An apparatus has been developed, called the electro-analogue, in which are incorporated all the elements required for building up any desired model of the process and of the automatic controller, as also an oscilloscope, a generator supplying voltage pulses ("step functions of voltage") and other accessories. Further articles will be devoted to the description of this electro-analogue.

Once the parameters of the model of the automatic controller have been chosen for optimum regulating conditions, then — taking into account the ratio of the time scales and the sensitivities — it is possible to derive the values of the settings in the actual regulating device.

Another method of determining the behaviour of a process is to plot the frequency response curve. To this end, the input quantity is caused to fluctuate sinusoidally with a constant amplitude about a normal value, at different frequencies. Measurements are then taken of the amplitude of the output quantity (in relation to the amplitude of the input) and of the phase shifts taking place between the fluctuations of the input and the output quantities. These two measured quantities can in our case most advantageously be worked out in a polar diagram. With this curve, together with the corresponding curve for the automatic controller, the stability of the regulating system can be judged.

The step-function response curve gives theoretically just as much information as the frequency response curve, with the advantage that it can be recorded much quicker and more easily than the frequency response curve. Taking again the example of a gas-fired furnace, the step-function response curve is obtained by suddenly opening the gas valve a little wider or turning it back a little and then measuring from that moment onwards the temperature inside the oven as a function of time. To record a frequency response curve it would be necessary to vary the gas supply sinusoidally and at a variable frequency and then, after a stationary state has been reached, the amplitude of the temperature fluctuation and the phase difference with the fluctuating gas supply have to be measured.

A second advantage of the method of working with step-function response curves is the simple way of designing the model in which the behaviour of the process is to be imitated. When the step-function response curve of this model is displayed on the screen of an oscilloscope it is not difficult to make it equivalent to that of the process by varying some of the network parameters. It would not be so easy to find agreement of the frequency response curves by oscilloscopic means.

Against these advantages of the step-function response curve, however, is the objection that disturbances occurring in actual processes are not usually of the nature of an impulse but mostly more or less gradual and periodical, so that the plotting of the frequency response curve is more in line with actual conditions. Furthermore, natural oscillations of the process often find better expression in the frequency response curve than in the step-function response curve.

While plotting either of these curves in the process it is necessary to see that all other working conditions — in the case of a furnace, for instance, the charge, the ambient temperature, etc. — are kept constant.

In the foregoing it has been tacitly assumed that the amplitude of the input disturbance or the magnitude of the step function is immaterial, and that the process is therefore linear. Often, however, it is only in the case of small variations that the process can be regarded as linear with a sufficient approximation. A complete investigation would have to cover also the behaviour of the process in the event of larger amplitudes, and the non-linearity of the actual process would have to be imitated as closely as possible by using non-linear elements in the network.

Any deviation from the linearity can be judged firstly by varying sinusoidally the input quantity of the process and observing in how far the output quantity is non-sinusoidal, and secondly by determining in how far non-linearity exists between input disturbance and output for disturbances of different magnitude.

In these considerations, however, we shall disregard non-linear effects.

Finally it is possible, at least in theory, to treat the problem of regulating mathematically. To facilitate the formulation of the necessary equations the letter Q is taken to represent the position of the regulating unit and the letter X for the deflection of the measuring element indicating the regulated quantity. This will prove to be very practical also when recording the step-function or the frequency response curve.

The fact that this definition differs somewhat from that given for fig. 2 is understandable when one bears in mind the example of the gas-fired furnace. Owing to the time delay of the detecting element the reading of the measuring element (now denoted by X) will not entirely agree with the temperature prevailing at the very same moment, originally denoted by X, and the flow of gas (Q in fig. 2) depends, it is true, upon the position of the gas valve now denoted by Q, but it also depends upon such factors as the pressure in the supply line. In the new definition the detecting element and regulating unit are reckoned as belonging to the plant.

In the relation between Q and X describing the process there will, as a rule, be also differential quotients with respect to the time t:

$$\mathrm{f}_1\left(Q, \frac{\mathrm{d}Q}{\mathrm{d}t}, \frac{\mathrm{d}^2Q}{\mathrm{d}t^2}, \ldots, X, \frac{\mathrm{d}X}{\mathrm{d}t}, \frac{\mathrm{d}^2X}{\mathrm{d}t^2}, \ldots\right) = 0. \quad (1)$$

(When Q is given a step-function or a sinusoidal variation with time, then it is possible to derive from this equation the step-function response curve or the frequency response curve as the case may be.) For the automatic controller, too, a relation between Q and X can be formulated:

$$\mathrm{f}_2\left(Q, \frac{\mathrm{d}Q}{\mathrm{d}t}, \frac{\mathrm{d}^2Q}{\mathrm{d}t^2}, \ldots, X, \frac{\mathrm{d}X}{\mathrm{d}t}, \frac{\mathrm{d}^2X}{\mathrm{d}t^2}, \ldots\right) = 0. \quad (2)$$

When the plant and the automatic controller are joined together to form one closed circuit (fig. 2b) then equations (1) and (2) may be combined. From these equations and the initial conditions one can, at least in theory, solve X and Q, which are then found as functions of time:

$$X = \mathrm{f}_3(t)\,, \quad \ldots\ldots\ldots \quad (3)$$

$$Q = \mathrm{f}_4(t)\,. \quad \ldots\ldots\ldots \quad (4)$$

From the curve representing (3) it can be ascertained whether the regulation is stable and whether it is sufficiently accurate, thus, e.g., whether X remains within the limits set or if, in the event of it exceeding those limits, it returns within them quick enough. From the graph representing (4) it can be seen whether the necessary variation of Q falls within the range covered by the regulating unit.

Actually, however, it is not as a rule an easy matter to solve X and Q from the equations (1) and (2). It is then the electrical model that provides the solution: with this model it is easy to observe the variations of the output and input voltages with time and to ascertain the effect of changes of certain parameters, introducing a disturbance at a suitable point in the circuit, either a step or a sinusoidal disturbance. In the latter case it is preferable to choose a frequency corresponding to the frequency of the disturbances occurring in practice, obviously taking into account the ratio of the time scales.

The quality of a regulating system can be judged from various points of view. Sometimes it suffices to consider the maximum deviation in the output quantity caused by a disturbance of a certain magnitude introduced somewhere in the system. In other cases the time taken for the output quantity to be brought back and kept within certain limits after the occurrence of a certain disturbance is more important; see *fig. 3a*. If a small deviation of long duration is in principle just as undesirable as a large deviation of short duration, then it may be taken as a criterion, for instance, that the hatched area in fig. 3b,

$$\int_0^\infty |X - X_0|\,\mathrm{d}t\,,$$

in which X_0 represents the desired value of X, or the integral

$$\int_0^\infty (X - X_0)^2\,\mathrm{d}t\,,$$

has to be kept below a certain value. No appreciable error is made if the integration is extended only over a time T of such a length that X becomes practically constant within that time; the values of the integrals extended over T can then easily be measured electrically in the model, the first value with the aid of a rectifier and a moving-coil meter and the second with a thermocouple and a moving-coil meter.

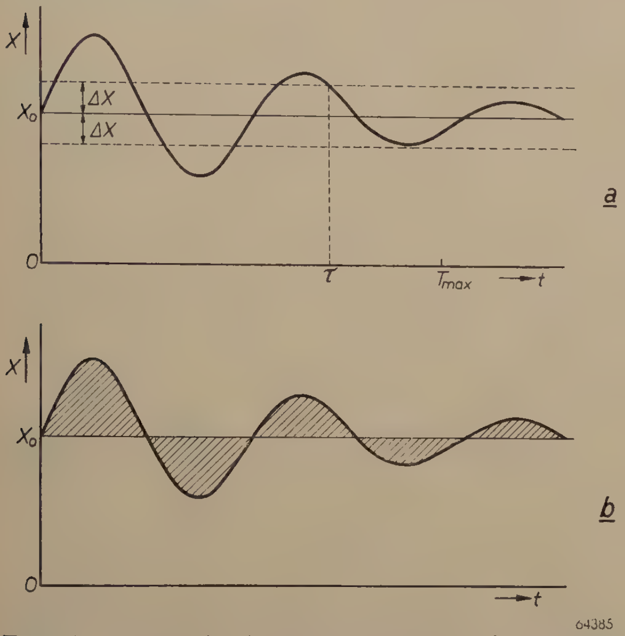

Fig. 3. Two criteria for the potential correction of a regulating system. In the case (a) the criterion is that the time t required for the quantity X to return permanently within the limits X_0—X and X_0+X must be less than a given time T_{max}. In the case (b) it is required that the hatched area does not exceed a certain value.

Types of regulating systems

Equation (2) expresses in what way the output of the controller depends on the input quantity. In principle one is quite free in the choice of this function, but in practice one is limited to functions that are easily attainable with mechanical or electrical switching elements. The most common combinations are mentioned below.

These are to be divided into two main groups: continuously and discontinuously acting controllers. The latter being the simpler in construction, for which reason they were first preferred in the technique of regulation (temperature control of furnaces, of centrally heated buildings, and suchlike), these will be dealt with first.

Discontinuously acting controllers

A controller is said to be discontinuously acting when the regulating unit takes up a different position discontinuously upon X passing a given critical value.

We shall first consider the case where all working conditions are normal. Let Q_0 be the value of Q necessary to give X the desired value X_0. In the simplest case the regulating unit is so arranged that the position Q_0 is not possible, there being only two possible positions: Q_1 ($< Q_0$) and Q_2 ($> Q_0$). According to the process equation, Q_2 corresponds to a value of X less than X_{cr}, while Q_2 applies for a value of X greater than X_{cr}. The controller reacts to the magnitude of X in such a way that either $X > X_{cr}$, $Q = Q_1$, or $X < X_{cr}$, $Q = Q_2$. One then speaks of a regulating system with two-step action (*fig. 4*). Naturally such a system is always unstable, for when X, for instance, is too great ($X > X_{cr}$) the regulating unit assumes the position $Q_1 < Q_0$ and Q therefore becomes too small, so that after a time X also becomes too small, with the result that the regulating unit again takes up the position Q_2 and X becomes greater again, and so on. Thus X fluctuates continuously about the value X_{cr} and the regulating unit alternates between the positions Q_1 and Q_2. Generally one will try to make the mean value of X equal the desired value X_0. In a certain working position X_{cr} may be equal to X_0 (see the part of *fig. 5* between $t = 0$ and $t = t_0$).

Let us now take a case where one of the working conditions changes, or where the controller is so adjusted that the regulating unit changes position at a different critical value X_{cr}' (see fig. 5 to the right of t_0). The only manner in which the controller can respond to the new situation is by changing the ratio of the intervals in which the regulating unit takes up the two possible positions. Since X_{cr}' is assumed to be greater than the original value X_{cr} the regulating unit has to remain longer in the high position Q_2 than in the low position Q_1. The new value X_0' around which X now fluctuates is higher but, as the diagram shows, remains lower than the new adjustment of the controller, X_{cr}'.

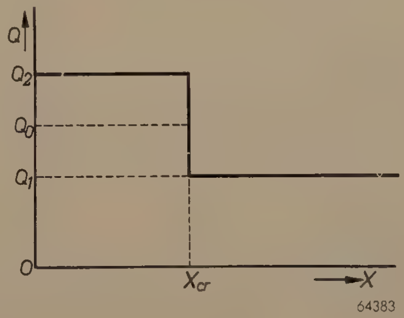

Fig. 4. Regulating system with two-step action. When X is greater than a certain critical value X_{cr} then Q has the value Q_1, and when X is less than X_{cr} then $Q = Q_2$.

A regulating system with two-step action can be extended to one with multi-step action working according to the following scheme:

when $X > X_2$ then $Q = Q_1$,
when $X_1 < X < X_2$ then $Q = Q_2$ $(>Q_1)$,
when $X < X_1$ then $Q = Q_3$ $(>Q_2)$.

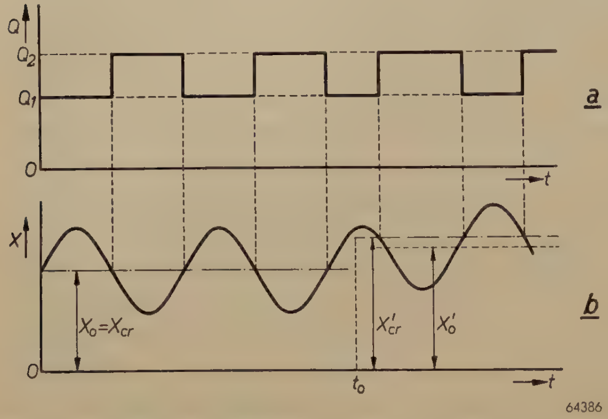

Fig. 5. Variations of Q and X in a regulating system with two-step action. Prior to t_0, X_{cr} was equal to the desired value X_0, around which X fluctuates, and Q had alternately, during equally long intervals, the values Q_1 and Q_0. At t_0 X_{cr} is raised to the value X_{cr}', as a consequence of which Q has each time the value Q_2 for a longer interval than the value Q_1. The value X_0' around which X then fluctuates is less than X_{cr}'.

Such a system has the advantage of reducing the variations of the input quantity and also the fluctuations of the process, so that the regulating unit remains at rest during longer intervals, thus being subject to less wear than in the case of a regulating system with two-step action.

Instead of causing the quantity Q itself to change at certain values of X, also the rate at which Q changes, thus $\mathrm{d}Q/\mathrm{d}t$, can be made to assume a different value. One then speaks of a multi-speed floating system. *Fig. 6* is a representation of the case where there are two speeds, p_1 and p_2. In practice there always is a certain zone (dead zone) between p_1 and p_2 where the system is at rest:

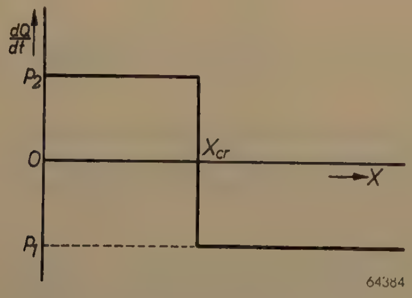

Fig. 6. Multi-speed floating system. Upon X passing the critical value X_{cr} the speed $\mathrm{d}Q/\mathrm{d}t$ with which the regulating unit is adjusted suddenly changes from the value p_1 to the value p_2, or vice versa. p_1 and p_2 have opposite signs.

when $X > X_2$ then $\mathrm{d}Q/\mathrm{d}t = p_1 < 0$,
when $X_1 < X < X_2$ then $\mathrm{d}Q/\mathrm{d}t = 0$,
when $X < X_1$ then $\mathrm{d}Q/\mathrm{d}t = p_2 > 0$.

This system of regulation is a transition to that of continuously acting regulating systems, since only the speed changes discontinuously, whereas the position of the regulating unit changes continuously.

Continuously acting controllers

The simplest form of a continuously acting regulating system is that with proportional action where equation (2) takes the simple form of:

$$Q = a(X - X_0) + Q_0, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

in which a is a constant. Q_0 is the position of the regulating unit for a stationary state of $X = X_0$ (eq. (1)).

In the event of a disturbance, caused by a change in the process equation, as a rule Q will have to assume a different value if the process is again to be brought to the state where $X = X_0$. Q and $a(X - X_0)$ — if Q_0 is left unchanged — can only vary by the same amount, and thus in the new position X will differ from X_0. This difference is the offsett of the regulating system. The greater the value of a, the sensitivity of the controller, the less is this difference, but the increasing of a is limited by the stability in the same way as the amplification of an amplifier with positive feedback: when a certain value is exceeded oscillations arise, that is to say, the system becomes unstable.

This offset is avoided in a controller with integral action. Between Q and X we then have for the controller an equation of the form:

$$Q = b\int(X - X_0)\,\mathrm{d}t + Q_0\,, \quad . \quad . \quad . \quad (6a)$$

hence

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = b\,(X - X_0)\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6b)$$

in which b is a constant. In this case, therefore, the speed at which the regulating unit is adjusted is proportional to the deviation of the position of the measuring element from the desired value X_0. This system is therefore also called a proportional-speed floating system. So long as this deviation is not reduced to zero, the factor $\int(X - X_0)\,\mathrm{d}t$ continues to increase, the regulating unit thus continuing to move in the direction for correcting the deviation, until ultimately $X = X_0$. Consequently there is then no offset. The greater the value of b the quicker the final state $(X = X_0)$ is reached, but here again a limit is set to the increasing of this coefficient owing to instability setting in.

Combinations of controllers with proportional and integral action (compound action) are also possible:

$$Q = a(X - X_0) + b \int (X - X_0) \mathrm{d}t + Q_0,$$

and there are also combinations of each of these systems, or of both, with a third type: the controller with derivative action, where a term $c\,\mathrm{d}X/\mathrm{d}t$ has to be added to the relation between Q and X in the right-hand member. This constitutes a contribution towards the adjustment of the regulating unit which is proportional to the rate at which X departs from the desired value (or is approaching that value). As a rule it brings about a quicker approach to the final state, since a and/or b can be chosen larger without risk of the regulating system becoming unstable.

In principle it should be possible to construct regulating systems with a relation between Q and X containing also higher derivatives or multiple integrals of X, but for industrial purposes there seems to be little need of these in practice.

It is to be noted that integrating and differentiating can be carried out electrically to a very good approximation, as is in fact done in the electrical model of a continuously acting controller. In many cases it is advantageous to construct also the controller itself as an electrical apparatus, as will be shown in a subsequent article following the two next ones devoted to the electro-analogue, in which the functioning of different regulating systems will be explained with the aid of oscillograms.

Summary. In this introductory article on regulating systems for industrial processes emphasis is laid upon the offset and the dynamic inaccuracy, the stability and the damping, as being the most important points to be considered in the designing of a regulating device. Attention is drawn to the advantages of working with an electrical model of the process to be regulated and of the controller to be designed for it; with this model the phenomena can be observed with the aid of an oscilloscope. The controllers are classified under discontinuously and continuously acting controllers. In the latter it is possible to introduce, in different ratios, a proportional, an integral and a derivative action.

Further articles will be devoted to a discussion of the "electro-analogue" — a combination of instruments with which analogous models of any processes to be regulated, including the controllers, can be built up — and a description of the actual construction of a controller.

PRECISE CALIBRATION OF TUNING FORKS

by C. C. J. ADDINK.

681.831.3:534.321.71:534.63

The introduction into the field of music of the new international standard of concert pitch based on a frequency of 440 c/s for the note A_3 in the middle octave instead of 435 c/s still involves difficulties owing to the lack of good standards, the tuning forks used in practice leaving much to be desired. Hence the greater interest now being shown in accurately calibrated tuning forks.

An international conference held in London in 1939 recommended that such bodies as the "Comité Consultatif International Téléphonique", the "International Electrotechnical Commission" and the "Union Internationale de Radiodiffusion", as also the various national standardisation committees, should fix the concert pitch at a frequency of 440 c/s for the note A_3 of the middle octave — instead of the 435 c/s internationally accepted at Vienna in 1885 — and that this should be adhered to as closely as possible in all musical performances[1]). In practice the usual tuning forks are, it is true, tuned higher than 435 c/s but often differ, in the positive or the negative sense, several c/s from the figure of 440 c/s. Since there was no standard for the frequency of 440 c/s in this laboratory a method has been developed, as the first step in this direction, for measuring audio frequencies direct (i.e. by counting the number of oscillations in a time interval) and with extreme accuracy. With the aid of an apparatus working according to this method not only have tuning forks of 440 c/s been calibrated but also sets of 13 tuning forks, forming a tempered chromatic scale of one octave (*fig. 1*). The method referred to will be described here, but the question to be considered first is the degree of accuracy to be demanded from a tuning fork.

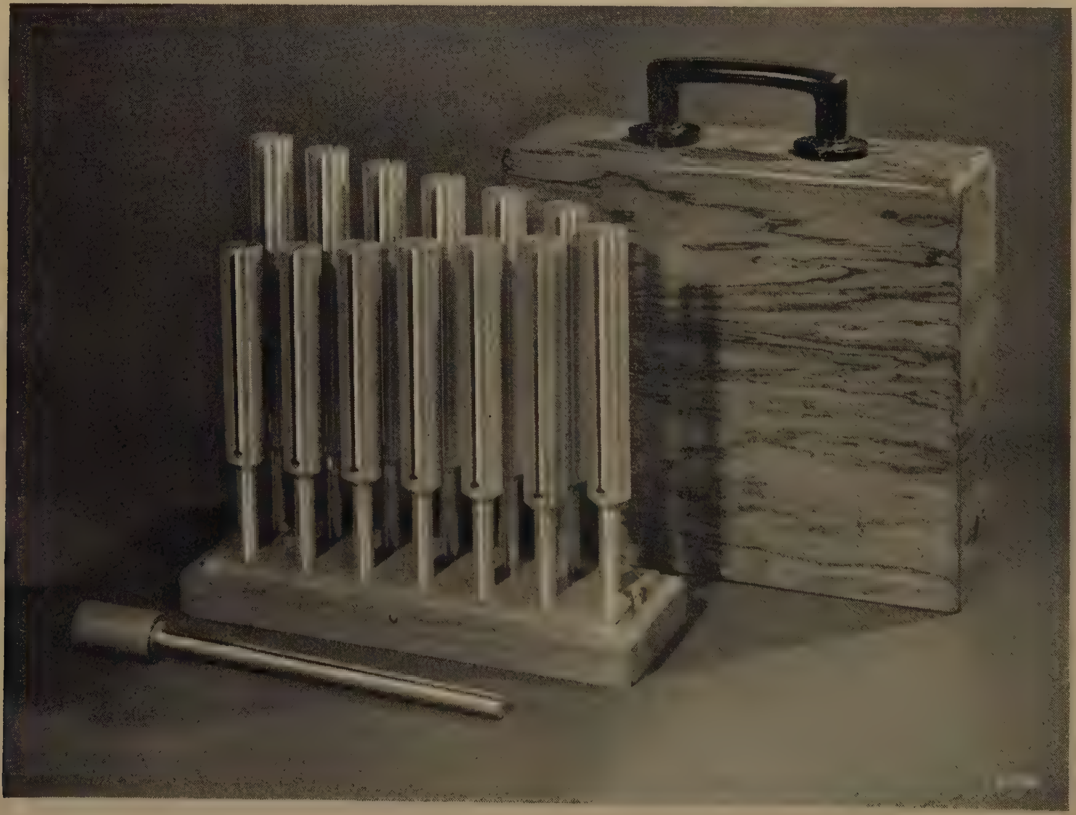

Fig. 1. Set of 13 tuning forks made in the Philips Laboratory. Together they form a tempered chromatic scale from c^1 to c^2, with $a^1 = 440$ c/s. The reason for the cylindrical shape is explained in the last paragraph.

---

[1]) See, e.g., Internationale Neufestsetzung des Stimmtones, Akustische Zeitschrift 4, 288, 1939.

Accuracy required of a tuning fork

The useful decay time of a normal tuning fork is 5 to 15 seconds. When the pitch of a musical instrument is compared with the tuning fork in the normal way one can usually hear beats. If, however, the difference in frequency is so small that there are no more than two or three beats during the decay time then they are hardly noticeable as such and one has reached the limit of the accuracy of the tuning fork. It serves no purpose, therefore, to require any greater precision than 0.2 c/s (one beat in 5 seconds), or in other words the frequency of the tuning fork has to lie between the limits of 439.8 and 440.2 c/s [2]). (This degree of accuracy is obviously only required for musical instruments with a fixed tuning, such as the organ, the piano, the harp, etc.) The frequency, however, has to be determined with greater accuracy, because one must be quite sure of the last decimal in the said limits. The permissible measuring error is therefore 0.05 c/s, which at a frequency of 440 c/s amounts to a measuring accuracy of about 0.01%.

This implies that account must be taken of the temperature when measuring and when using the tuning fork in practice. The temperature coefficient of the frequency in the case of a steel tuning fork is in the order of -1×10^{-4} per °C, while for the forks made of a hard aluminium alloy illustrated in fig. 1 it is about -2.5×10^{-4} per °C.

Method of measuring

As is evident from the foregoing, for calibrating a tuning fork a more accurate method is needed than that where the tuning fork is compared with a standard frequency by ear. Furthermore, with frequency differences less than 0.1% the sign of the error can no longer be determined by ear.

The method of calibrating worked out by us briefly amounts to the following. The frequency of an auxiliary oscillator is matched with that of the tuning fork to be calibrated and then a synchronous clock is connected to the oscillator and the oscillations made by the auxiliary oscillator in a certain interval of time are counted. This time interval is measured with the aid of a calibrated pendulum timepiece.

Matching the auxiliary oscillator

Owing to the damped character of the vibration a tuning fork is not suitable for measuring the frequency with sufficient accuracy without some other aids. It is true that a tuning fork can be kept in vibration by electrical means, but then, unless a great deal of care is taken, the frequency at which it vibrates differs somewhat from the frequency of a fork vibrating freely, and it is just this free vibration that has to be measured accurately. What is needed for calibrating, therefore, is an auxiliary oscillator which, once it has been matched with the freely vibrating fork, continues to oscillate in that frequency for an unlimited length of time.

An *RC* oscillator consisting only of resistors, capacitors and amplifying valves is very suitable for this purpose. If such an oscillator is properly designed, when it has reached temperature equilibrium the relative frequency change in the audio-frequency range can be kept smaller than $1:10^5$ [3]).

The matching of the frequency of the oscillator with that of the tuning fork can be done with sufficient accuracy if a cathode-ray oscilloscope is employed, by applying the signal from the *RC* oscillator to one pair of plates and a signal with the frequency of the tuning fork to the other pair. The latter signal can be obtained by setting up the tuning fork in front of a microphone. There then appears on the screen of the oscilloscope a changing Lissajous figure which, when the frequencies are equal, becomes a stationary ellipse [4]).

There is, however, a drawback to this method. The Lissajous figure cannot demonstrate whether the frequency deviation is positive or negative, hence the operator does not know in which direction to turn the oscillator tuning knob. If this is turned in the wrong direction the consequent time loss results in the tuning fork signal having decayed too far, and the fork has to be struck again.

To obviate this, the following method is adopted. The signal from an *RC* oscillator is applied direct to one pair of input terminals of the oscilloscope and also, after having been shifted 90° in phase in a network of resistors and capacitors, to the other pair of terminals (*fig. 2*). Given the right proportions of amplitude of the two signals, a circle is then produced on the screen of the oscilloscope.

The tuning fork to be calibrated is set up front in of a microphone, the amplified signal from which is used for modulating the beam current of the cathode-ray tube, such that the beam current is allowed to pass only during the positive half cycles of the microphone signal. Thus only half a circle is seen on the screen (fig. 2). When the oscillator frequency is exactly equal to the frequency of the tuning fork this half circle is stationary, but when there is a difference in frequency it rotates. The best method is to arrange the apparatus so that the tuning knob of the oscillator has to be turned clockwise to increase the frequency and that the half circle rotates clockwise if the oscillator frequency is too high and anti-clockwise if it is too low. Matching is then done very quickly, the vernier control of the oscillator being turned in the direction opposite to that in which the half circle is moving, until the latter is stationary [5]). (A half circle has been chosen here for the sake of simplicity, but this is not essential for matching the frequencies. One can also work with the half of any ellipse, since the phase shift in the network need not be precisely 90°.)

---

[2]) Other authors arrive at the same tolerance another way: H. J. von Braunmühl and O. Schubert, Ein neuer elektrischer Stimmtongeber für 440 Hz, Akustische Zeitschrift **6**, 299-303, 1941.

[3]) See, e.g., E. L. Ginzton and L. M. Hollingworth, Phase-shift oscillators, Proc. Inst. Rad. Engrs **29**, 43-49, 1941; K. Bucher, *RC*-Generatoren, Telegr. Fernsprech-Techn. **31**, 307-313, 1942.

[4]) For this and the other method to be described later it is an advantage to use an oscilloscope with two amplifiers, one for each pair of plates, such as type GM 3159 or GM 5655 (Philips Techn. Rev. **9**, 202-210, 1947, and **11**, 111-115, 1949 (No. 4)).

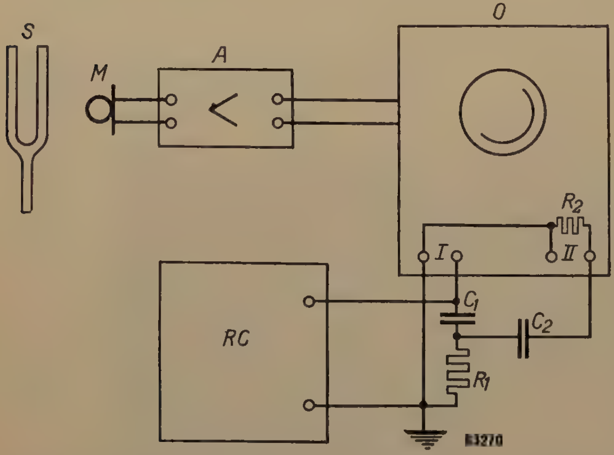

Fig. 2. Circuit for matching the frequency of an *RC* oscillator (*RC*) to that of the tuning fork to be calibrated (*S*). The signal from the oscillator is applied direct to the input terminals *I* of the cathode-ray oscilloscope *O* and, via a 90° phase-shifting network of resistors and capacitors (R_1, C_1, R_2, C_2), to the input terminals *II*. The beam current of the cathode ray tube is modulated by the output from an amplifier (*A*) to which is connected a microphone (*M*) picking up the sound from the tuning fork.

As indicated in the diagram, R_2 may be the resistor already contained in the oscilloscope between the terminals *II*.

If a voltage source with the standard frequency is available (e.g. an electrically driven tuning fork, or the 440 c/s signal as regularly transmitted by the B.B.C. for tuning purposes), then the tuning-fork frequency that is to be determined can be compared directly with the standard frequency by means of the set-up shown in fig. 2, replacing the *RC* oscillator by the standard frequency signal. When this method is followed a rotating half-circle is again seen, and from the number of revolutions it makes in a certain time it is easy to derive the difference in frequency (the sign is indicated by the direction of rotation). Thus one measures, as it were, a fraction of a beat. The accuracy is proportional to the duration of the observation, being already very great at the duration of the decay time. If, for instance, the half circle makes $^1/_4$ revolution in 12.5 seconds then the frequency difference is 0.02 c/s.

Measuring the oscillator frequency with the aid of a clock

A synchronous clock with a seconds hand is then connected to the *RC* oscillator matched to the tuning fork. In our case a clock was used running at the right speed when connected to a signal of 1000 c/s. If it is fed with a frequency f during a period of time T_1 and in that time its reading changes by an amount T_2 then

$$f = \frac{T_2}{T_1} \cdot 1000 \text{ Hz.}$$

T_1 is measured with a good pendulum timepiece checked before and after the measurement with radio time signals, which thus form the basis of the calibration. Since the seconds hand of this clock moves forward 1 second at a time, and in fact the two clock readings cannot be taken simultaneously with sufficient accuracy, a stop-watch is used, this being started when the seconds hand of the pendulum timepiece arrives at a certain dial mark, say 60. From that moment onwards the watch forms, as it were, an extension piece of the pendulum clock giving readings accurate to within 0.2 or 0.1 second according to whether there are 5 or 10 balance movements per second. If the watch is stopped when the continuously rotating seconds hand of the synchronous clock passes a certain dial mark then the moment that this has taken place is found by adding the position of the stopped watch to the time shown by the pendulum timepiece. The three readings (of the pendulum clock, the synchronous clock and the stop-watch) are noted down.

After a sufficient length of time the procedure is repeated and from the readings taken one can find immediately T_1 and T_2 (see the figures given below by way of example).

The error in the movement of the pendulum clock could be made smaller than one second per

---

[5]) This method has been used before, for measuring the pitch of musical instruments during a musical performance (Balth. van der Pol and C. C. J. Addink, Philips Techn. Rev. 4, 205-210, 1939. and Wireless World 44, 441-442, 1939). The oscillator — in that case a calibrated string with electrically sustained vibration and variable through a small frequency range — could thereby, within one second, be made equal in frequency to the note a^1 in the music to within 0.2 c/s. For calibrating a tuning fork, however, a string variable in frequency is not accurate enough.

24 hours, i.e. less than $1/10^5$, so that no correction need be made for this. Neither is any correction necessary for the movement of the stop-watch, since for the purpose of this test it runs for such a short time that any error can be ignored.

The reading errors at the beginning and at the end of the time T_1 being additive, when a stop-watch is used that can be read accurately to within 0.2 s the duration of T_1 is known accurately to within 0.4 s. Since, as deduced above, the measurement with a tuning fork of about 440 c/s has to be accurate within about 0.01%, the duration of T_1 has to be at least $0.4/10^{-4} = 4000$ s, thus about one hour. (With a stop-watch recording time to within 0.1 s the duration of T_1 need be only half an hour.)

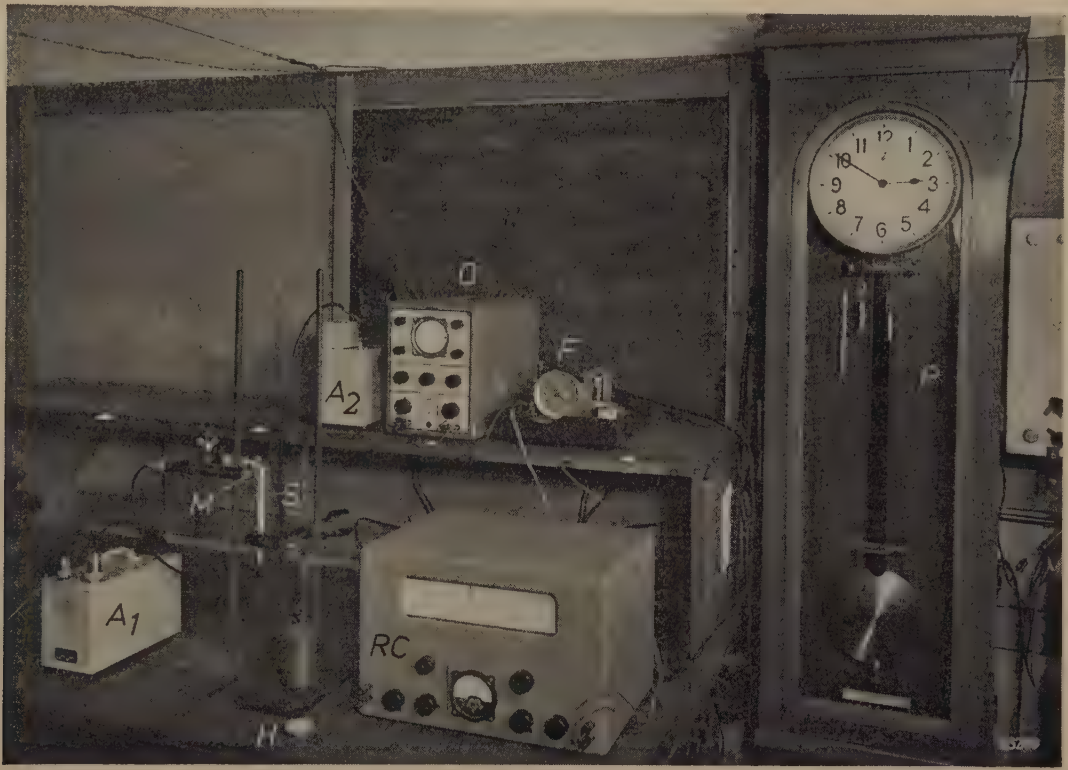

Fig. 3. Complete apparatus for calibrating tuning forks. *S* the tuning fork to be calibrated, *M* microphone, A_1 and A_2 amplifiers together forming the amplifier *A* of fig. 2, *O* oscilloscope, *RC* auxiliary oscillator, *P* pendulum timepiece, *F* synchronous clock, *H* stop-watch.

A picture of the complete calibrating apparatus is given in *fig. 3.*

Some readings taken during a practical calibration are given below.

The watch was started when the pendulum clock showed 2 hrs 37′ 0″ and stopped when the seconds hand of the synchronous clock passed a given mark of the dial, viz. when this clock showed 3 hrs 55′ 35″. The stop-watch then showed 7.2″. Thus at the moment that the synchronous clock showed 3 hrs 55′35″ (the beginning of T_2) the time shown by the pendulum clock was 2 hrs 37′7.2″ (the beginning of T_1). After about one hour, at 3 hrs 31′0″ on the pendulum clock, the stop-watch was started again and then stopped when the synchronous clock showed 4 hrs 19′20″ (end of T_2). The stop-watch then showed 7.8″ and was thus stopped at the time 3hrs 31′7.8″ on the pendulum clock (end of T_1).

Thus we find: T_1 = 3 hrs 31′7.8″ — 2 hrs 37′7.2″ = 3240.6″, and T_2 = 4 hrs 19′20″ — 3 hrs 55′35″ = 1425″. Thus $f = (1425/3240.6) \times 1000$ c/s = 339.73 c/s.

To minimize any inaccuracy due to errors in reading, the start and the finish of T_1 and T_2 are repeated twice; of the three readings thus obtained the identical ones are usually correct.

Correction of the tuning fork

If the tuning fork tested is found to deviate from the desired frequency then one will want to correct it. If the frequency is too low this can be done easily; the exact frequency can be approached very closely by removing a little material from the end of the prongs of the fork. This has to be done in such a way however, that no difference in the lengths of the prongs is introduced, for the slightest difference considerably increases the damping. For this reason, bifurcated rotationally cylindrical tuning forks were adopted (fig. 1). The prongs can be shortened as required by stopping the saw cut temporarily with, for example, a strip of brass, and then shaving the end face in a lathe.

Once the frequency f of the tuning fork has been measured, then using the formula giving the frequency as a function of the dimensions and material constants, an accurate calculation of the requisite modification can be made. The formula is [6]):

$$f = \frac{K^2 v}{(l + l_0)^2},$$

where K^2 is a factor dependent upon the shape of the tuning fork, v is the velocity of sound in the material of the fork, l the length of the prongs and l_0 a constant, which in the present case is small compared with l (a few %). The length of the prongs is taken to be the distance from the vibration node to the tip; with the form described above the vibration node lies on one level with the centre of the bore at the bottom of the saw cut.

After the prongs of a fork having too low a frequency have been shortened by the amount calculated according to this formula, the remaining frequency deviation is usually less than the tolerance allowed (± 0.2 c/s).

If the tuning fork has a frequency slightly too high it is still possible to correct it, by making the prongs a little thinner at the level of the vibration node. It can hardly be calculated in advance, however, how much material has to be removed, and furthermore the appearance of the tuning fork is thereby spoiled. For this reason the uncorrected tuning forks are always given a frequency slightly too low.

This type of fork is made in a lathe and a milling machine. A rod is turned to produce the required diameters for the prongs and the stem, and symmetry is assured. A hole is now drilled at a right angle through the axis of the rod. A slot, in line with the drilled hole, is cut along the axis of the rod on a milling machine. It is obvious that this is a much simpler process than that of making the usual rectangular cross section fork.

So as to have a light and stainless product a hard aluminium alloy was chosen instead of steel, although it has a somewhat greater absolute temperature coefficient.

A disadvantage due to the shape and the material chosen is that the fork has a weaker tone and a shorter vibration time than the usual fork when placed on a sounding board. However, this fork gives better results when held close to the ear, especially at low frequencies, owing to the relatively larger radiating surface of the prongs.

Summary. A method is described by means of which tuning forks can be calibrated without requiring an acoustic standard frequency. With the aid of a cathode-ray oscilloscope the frequency of an *RC* oscillator is matched to that of the tuning fork to be calibrated. Connected to this oscillator is a synchronous clock, the seconds hand of which would make one revolution per minute if the frequency were 1000 c/s. If, after an interval T_1, this clock shows a difference in reading T_2 then the frequency sought is $f = (T_2/T_1) \cdot 1000$ c/s. The duration of the interval T_1 is measured with the aid of a pendulum timepiece the movement of which is checked before and after the measurement with radio time signals.

If the frequency found is lower than that required then the latter can be very closely approximated by shortening slightly the prongs of the tuning fork. A special, cylindrically shaped tuning fork is described, with which this correction can be made easily on a lathe, thus retaining the required symmetry.

---

[6]) See, e.g., Handbuch der Physik, part. 8, p. 207 (Springer, Berlin 1927).

EXPERIMENTAL TESTING OF ELECTRICAL NETWORKS BY MEANS OF THE UNIT FUNCTION RESPONSE

by J. van SLOOTEN.

621.317.75:621.317.333.8

As is known, the response of an impedance or an electrical network to the so-called unit function current or voltage gives valuable information about the nature of the impedance or the network. Hitherto, however, very little use has been made of this for answering questions which repeatedly occur in experimental work in a radio laboratory, such as: How great is the inductance of this coil? What is the damping of that circuit? Is this filter terminated with the right resistance? A simple measuring device, using the method of unit function response, has been developed in order to meet this need.

Unit function response curve

In the theory of electrical networks particular attention has long been paid to the reaction of a network to the sudden switching on or off of a direct voltage or a direct current at the input of the network, i.e. to the response to a voltage or a current in the form of a so-called unit function. The graphical representation of this response as a function of time is called a unit function response curve. It is known that this curve tells us just as much as the frequency response curves describing the amplitude and phase of the response of the network as a function of the frequency [1]). Mathematically speaking, the unit function response curve is therefore equivalent to the frequency response curves. This does not mean to say, however, that in practice it is immaterial which of these curves is available. In the television and radar techniques one frequently has to deal with pulses of finite duration and it has for long been known that in such cases the unit function response curve is more valuable than the frequency response curves: the reaction of a network to a pulse of finite duration is found by taking the difference between two equal unit function response curves, one of which is shifted over the width of the impulse with respect to the other.

Complicated instruments for studying the response of networks to pulses of short duration have already been developed [2]). In very many cases, however, a much simpler apparatus will be able to render just as good service, not only for studying fairly complex networks but also for all kinds of routine measurements. Particularly if no stringent requirements are imposed as regards the accuracy of the measurements — for example in the case of sampling tests in the manufacture of radio parts like coils, I.F. tranformers, etc. — much time can be saved by employing a simple measuring device as described below.

By means of this simple device the response of an impedance or a network to a current in the form of a unit function is traced on the screen of a cathode-ray tube. From the oscillogram thus obtained it is possible to draw various conclusions, as some examples will show. The fact that the equipment can be so simple is due to the use of a unit function instead of pulses. The response to a pulse, as already remarked, can easily be deduced from the oscillogram if such is desired.

When the response of a network to a current impulse (of finite duration) is compared with the response to a unit function of current of the same amplitude, these responses are only of the same order if the duration of the pulse is of the order of one cycle of the response (if this is periodical) or of the time constant of the response (if it is non-recurring). This implies that in an apparatus working with impulses both the interval between the impulses and the duration of the impulses have to be separately variable. In principle, therefore, such an apparatus requires a more complicated system than one working with unit functions of current.

Description of the apparatus

The apparatus consists of a device supplying the unit function of current to be sent through the impedance or network under test, and a cathode-ray tube with its accessories.

Since the response has to be investigated by means of an oscillogram, the unit function has to be repeated periodically, and for this purpose use is made of the well-known Abraham and Bloch multivibrator. This multivibrator can be so designed as to furnish not only a rectangular voltage with the desired frequency — which voltage controls a valve supplying the rectangular current — but also a sawtooth voltage required for the horizontal deflection of the electron beam in the cathode-ray tube.

---

[1]) See, e.g., J. Haantjes, Judging an amplifier by means of the transient characteristic, Philips Techn. Rev. **6**, 193-201, 1941.

[2]) See, e.g., D. S. Espley, E. C. Cherry and M. M. Levy, The pulse-testing of wide-band networks, J. Inst. El. Engrs. **93** III A, 1176-1187, 1946.

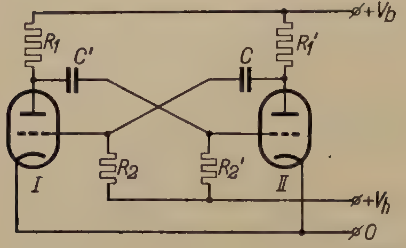

Fig. 1. Basic circuit of the multivibrator. *I* and *II* triodes (or pentodes), R_1 and R_1' anode resistors, R_2 and R_2' grid resistors, C' and C coupling capacitors, V_b supply voltage, V_h auxiliary voltage.

The circuit diagram of the multivibrator is shown in *fig. 1*. The two valves *I* and *II* are interconnected in such a way that the voltage drop in the anode circuit of one valve keeps the other valve cut off, and vice versa. As soon as some current begins to flow through the valve first cut off (owing to decrease of the negative grid voltage), say *I*, the positive feedback between the two valves causes the anode current in valve *I* to increase further. As a consequence the anode voltage of valve *I* drops, the grid of the valve *II* becomes more negative and the anode current in valve *II* decreases until that valve is entirely cut off. This process is then repeated in the reverse direction, and so on. The transition of current from one valve to the other is so rapid that the variations of the anode voltage of the valves are approximately rectangular.

So long as a valve of the multivibrator is conducting its grid voltage is practically zero. While the valve is cut off its grid voltage changes according to an exponential function, but if the grid resistors are connected to a point of high positive voltage then that function is practically linear, because only the first, almost rectilinear, part of the exponential curve is traversed. The positive voltage referred to (V_h in fig. 1) may be equal, for instance, to the supply voltage V_b (300 V) of the anode circuits.

Fig. 2 represents the variations of the voltages on the anode and on the grid of one of the valves. The most important deviation from a rectangular form of the anode voltage is indicated by the dotted curves; at the moment that the valve (say *I*) is cut off the anode voltage does not immediately become equal to the supply voltage V_b. The reason for this is that in the other valve there is at first a fairly strong grid current, drawn mainly from the coupling capacitor (C' in fig. 1), causing a voltage drop in the anode resistor (R_1). This has the effect of making the shape of the anode voltage somewhat less rectangular at the moment the valve is cut off than at the instant when it becomes conductive. This sharp change in the voltage when the valve becomes conductive is therefore used to bring about a sudden drop of the anode current in a third valve, the switching valve, to zero. The input impedance of the network to be tested is incorporated in the anode circuit of the switching valve (a pentode) and in this way the unit function response (the voltage across the impedance or across the output terminals of the network) can be observed on the screen of the oscilloscope.

In principle there is no necessity for the third valve, because the impedance to be investigated could be directly connected in series with one of the anode resistors of the multivibrator, since unit function currents also occur in these resistors. Unless, however, the impedance is very small compared with the anode resistances, the working of the multivibrator is disturbed. Hence the desirability of using a separate, third valve as switching valve.

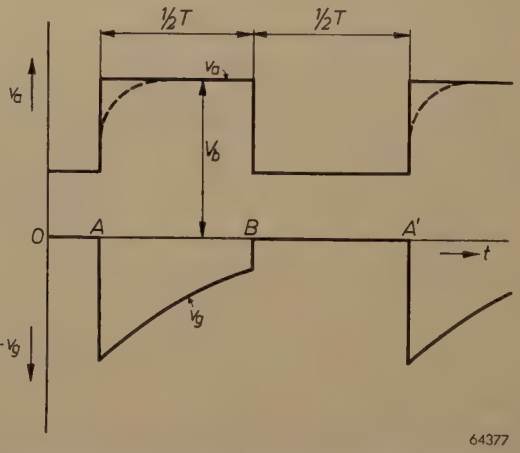

Fig. 2. Anode voltage v_a and grid voltage v_g of one of the multivibrator valves (on different scales). When the valve is cut off (at *A*) the dotted curve is obtained, which does not occur when the valve begins to pass current (at *B*).

The frequency of the multivibrator has to be such that the switching-off transient to be investigated almost completely dies out within half a cycle. Since the current flowing through the switching valve and the impedance under investigation changes alternately from zero to a certain value I_{max} and back again to zero, there is alternately a switching-on and a switching-off transient. In order to avoid intermixing of these two phenomena on the screen of the oscilloscope, the switching-on transient is either rendered entirely invisible or compressed to a vertical line at the side of the oscillogram; this will be reverted to later.

Fig. 3 is a complete circuit diagram of the oscilloscope with its accessories. The impedance to be investigated, *X*, is connected between the terminals *p* and *r*. With the switch S_2 in the position shown in the drawing the voltage across the impedance is applied to the plates for the vertical deflection. When a four-terminal network has to be investigated then the input terminals are connected to *p* and *r* and one of the output terminals to *q*; with S_2 in the other position the voltage variation at the output terminal is then depicted on the screen.

The direct-current resistance between the terminals *p* and *r* must not be higher than about 5000 ohms, in view of the maximum permissible voltage drop (150 V) caused by the D.C. anode current (30 mA); sufficient voltage has to be left for the proper functioning of the switching valve *III*. Further, it is necessary to make sure that when the switch S_2 is in the upper position there is always a conducting connection between the points *q* and *r*, so as to keep the average potential difference between these points at zero level; the resistance of this connection may be high, up to about 1 megohm.

The frequency of the multivibrator, and therefore the duration $T/2$ of a half cycle in which the oscillogram is described, is varied by switching on (with S_1) coupling capacitors of different capacitances. Suitable values for $T/2$ are, for example, 1000 μs and 40 μs. It is of importance to know this duration exactly, and the easiest way to determine it is to connect a resistor across the terminals *p* and *r* and then to compare on an oscilloscope the frequency $1/T$ of the rectangular voltage across that resistor with the frequency of a calibrated oscillator. Another (less accurate) method, dispensing with an oscillator, will be described later.

One of the plates for the horizontal deflection (see fig. 3) is connected to one of the grids of the multivibrator valves via a coupling capacitor. From fig. 2 it is seen that during one half cycle the grid

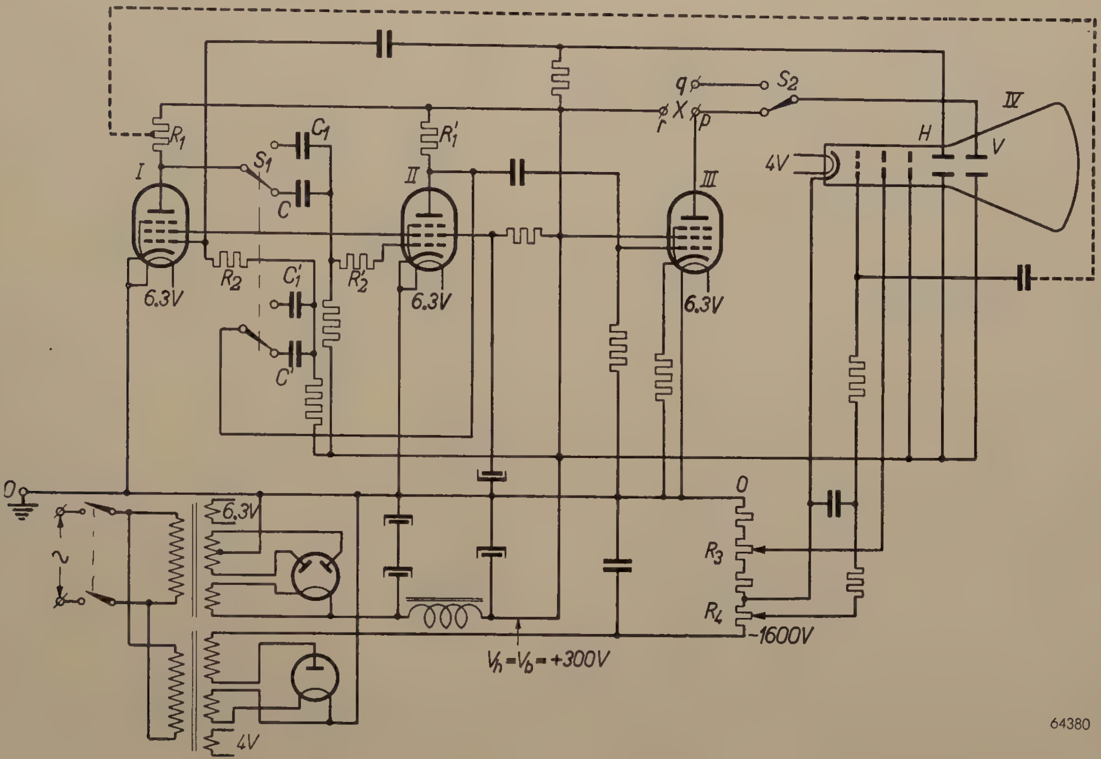

Fig. 3. Complete circuit of the oscilloscope. *I*, *II* multivibrator valves (type EL 41), *III* switching valve (EL 41), *IV* cathode-ray tube (DG 10-3), R_1 and R_1' anode resistors, R_2 and R_2' grid resistors of the multivibrator. Two different frequencies are obtained by connecting the coupling capacitors *C* and *C'* or C_1 and C_1' by means of the switch S_1. The terminals *p* and *r* are connected to the impedance *X* under investigation (switch S_2 in the position drawn) or to the input of a four-terminal network under investigation, the output of which is connected to *q* (S_2 in the upper position). *H* plates for the horizontal deflection, *V* plates for the vertical deflection. The electron beam is focused with R_3 and the brightness of the oscillogram is adjusted with R_4.

voltage rises almost linearly, so that a likewise practically linear time base is obtained, and that during the other half cycle the grid voltage is about zero. The latter means that when the switching-off transient of the impedance under investigation is produced on the screen, the picture of the switching-on transient is compressed to a single vertical line on the extreme right of the picture, where it is not disturbing (in many cases it is even useful, as will be shown presently with some examples). If desired the vertical line can be entirely suppressed by applying to the control grid of the cathode-ray tube, during the respective half cycle, a negative voltage which can be taken from the multivibrator by means of the connection indicated by the dotted line in fig. 3.

Some examples will now be given to show how the oscilloscope described here can be used to advantage.

Examples of application

Resistance

The simplest impedance to be investigated is of course a resistance, one of say 3000 ohms. The oscillogram produced will then have the appearance of that given in *fig. 4*. During the half cycle that the control grid voltage of the valve *I* (fig. 3) is about zero (*BA'*, fig. 2) the spot is somewhere on a vertical line. At the beginning of this half cycle the switching valve suddenly passes current and the resulting voltage drop in the resistor *R* to be investigated causes the spot to travel upward [3]). Thus the switching-on transient appears on the screen as a vertical line, the length of which is proportional to *R*. At the end of that half cycle (at *A'*, fig. 2) two things happen simultaneously: the current through the switching valve is suddenly cut off and the grid voltage of the valve *I* suddenly assumes a high negative value. As a consequence the spot of light very rapidly moves downward and across to the left, the exact course it follows being immaterial (in fig. 4 this is indicated by a lightly drawn line). In the half cycle then starting, the voltage across the resistor remains zero, while the grid voltage of the valve *I* gradually increases, the spot thus describing a horizontal line from left to right and returning to the starting point at the moment that the next switching-on transient sends it upward again, and so on [4]).

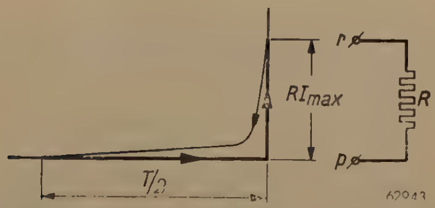

Fig. 4. Oscillogram obtained from a resistor *R* connected to the terminals *p* and *r* (fig. 3). The length of the vertical line on the right is proportional to the resistance *R*, the anode current amplitude $I_{\max}$ and the sensitivity of the cathode-ray tube.

The "sensitivity" of the oscilloscope can be determined by connecting a known resistance. It is in the order of 1 cm deflection per 1000 ohms. The length of the vertical line obtained with a certain resistance can be compared with the oscillogram of composite impedances.

Resistance and capacitance in parallel

When a capacitor *C* of comparatively low value is shunted across a resistor *R* ($RC \ll T/2$) an oscillogram is obtained similar to that shown in *fig. 5*. On the right is again the compressed picture of the switching-on transient. When the current in the switching valve has dropped from $I_{\max}$ to zero the voltage *v* across the *R*-*C* circuit decreases exponentially with the time *t* according to the function:

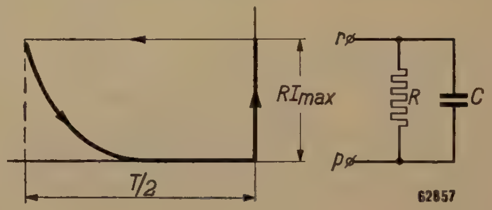

Fig. 5. Oscillogram obtained from a resistor *R* shunted by a small capacitor *C* ("small" means $RC \ll T/2$).

$$v = R\,I_{\max}\,\mathrm{e}^{-\quad RC} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

During the transition of current from one multivibrator valve to the other, the electron beam in the cathode-ray tube first moves rapidly from the top right-hand corner of the oscillogram to the top left-hand corner and then more slowly downward to the bottom right-hand corner. If the oscilloscope has been carefully mounted, so that stray capacitances are kept low, this rapid flyback takes no longer than about 1 μs and thus this interval of time is negligible compared with $T/2$.

---

[3]) Here and in the following examples it is assumed that the plates for the vertical deflection are connected in such a way that the voltage drop caused by the current in the impedance to be investigated results in an upward deflection.

[4]) With the process described here the oscillogram comes to lie roughly in the middle of the screen, since a blocking capacitor ensures that the mean value of the deflection voltage for the horizontal direction is zero.

If the duration of the stroke $T/2$ is known, it can be estimated how long it takes for the voltage v to drop to 1/e (about 37%) of the peak value. Since this time is equal to RC (see eq. (1)), when the value of R is known a rough estimate can be made of the capacitance C.

Conversely, with known values of R and C the duration of the stroke $T/2$ can be determined, without having to use the calibrated oscillator referred to above. The greatest accuracy is obtained if RC and $T/2$ are of the same order. The oscillogram then assumes the shape represented in *fig. 6*.

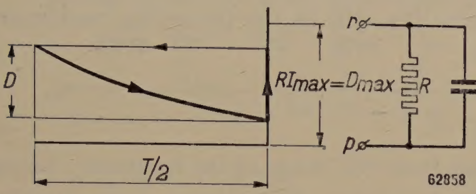

Fig. 6. Oscillogram obtained from a resistor R shunted by a capacitor C in the case where RC is not small with respect to $T/2$.

During the half cycle that the current is switched off the voltage v remains well above zero and while the current is switched on it does not reach the value $RI_{\max}$. A simple calculation shows that the vertical amplitude D on the screen then has the value:

$$D = D_{\max}\frac{1-p}{1+p}, \quad \ldots\ldots\ldots \quad (2)$$

where $p = e^{-T/2RC}$ and $D_{\max}$ is the amplitude corresponding to $RI_{\max}$ obtained when C is of a very low value. If C is so chosen that $D = \frac{1}{2}D_{\max}$ then from eq. (2) it follows that $p = {}^1/_3$, so that

$$\tfrac{1}{2}\,T = RC\log_e 3 = 1.1\,RC\,.$$

Thus, with R and C known, it is possible to determine the duration of the stroke $T/2$.

Resistance and self-inductance in parallel

Let us now consider the case where an inductance L is connected in parallel with a resistance R and the ratio L/R is small compared with $T/2$.

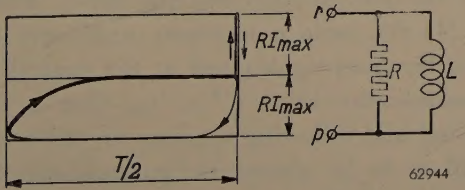

Fig. 7. Oscillogram obtained from a resistor R shunted by an inductor L in the case where $L/R \ll T/2$.

The oscillogram then gives a picture similar to that in *fig. 7*. Owing to the delay caused by the inductance the current continues to flow through L for some time after the current in the switching valve has dropped to zero. This gives rise to a positive voltage across R, and thus, under the conditions assumed in footnote [3]), a downward deflection, which decreases exponentially according to the formula:

$$v = RI_{\max}\,e^{-\frac{R}{L}t}\,.$$

In a manner similar to that described above for capacitances, inductances can thus be roughly measured, the most accurate results being obtained by comparison with one or more known inductances.

L-C circuit

The oscillogram obtained with an L-C circuit is that of a damped oscillation (*fig. 8*), provided the damping of the circuit is not so great as to cause the phenomenon to be aperiodic.

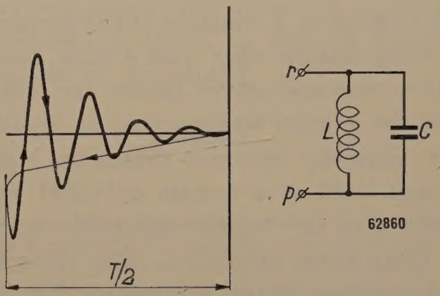

Fig. 8. Oscillogram obtained from a damped oscillatory circuit.

The formula for the voltage v across the capacitor is:

$$v = V_0\,e^{-\frac{r}{2L}t}\cos\omega_0 t,$$

where V_0 is the initial value of v, r is the loss resistance taken to be concentrated in the inductance, and $\omega_0 = 2\pi$ times the natural frequency of the circuit.

The damping of the circuit can easily be determined from the ratio of the successive peak values of v. Inductances can also be measured by comparing the number of oscillations in a certain interval of time with the number occurring with the same capacitance and a known inductance.

Band-pass filters (I.F. transformers)

The band-pass filters so frequently employed in radio technique can also be studied with the aid of the unit function response curve. A band-pass filter may consist of two L-C circuits coupled, for instance, by a mutual inductance M (see *fig. 9*, where for the sake of simplicity the two circuits are assumed to be identical).

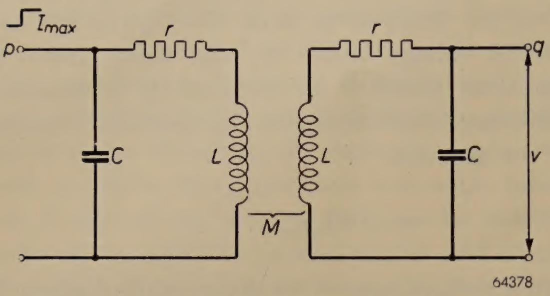

Fig. 9. Band-pass filter (I.F. transformer) with two identical circuits coupled by a mutual inductance M.

A characteristic of a band-pass filter is its coupling factor k, which , for the filter represented in fig. 9, is defined as M/L. When the behaviour of the filter is investigated as a function of k one finds a remarkable difference between the method of the unit function response and that of the frequency response. When the latter method is followed, thus measuring the amplitude of the output voltage as a function of the frequency, it is found, as is known, that with a loose coupling the frequency response curve shows one, rather low, peak at the angular frequency $\omega_0 = 1/\sqrt{LC}$, whilst as the coupling becomes gradually tighter the peak rises until at a certain critical coupling it cannot rise any higher and splits up into two peaks lying either side of ω_0. The value k_{cr} of the critical coupling factor is equal to the damping factor δ:

$$k_{cr} = \delta \equiv r/\omega_0 L = r\sqrt{\frac{C}{L}}\,.$$

For the variation (as a function of time) of the secondary voltage v of a band-pass filter, when a unit function of current $I_{\max}$ is applied to the primary terminals, the following approximate expression can be formulated:

$$v = I_{\max}\sqrt{\frac{L}{C}}\cdot e^{-\frac{r}{2L}t}\cdot \sin\tfrac{1}{2}\,k(\omega_0 t - \delta)\cdot\cos(\omega_0 t - \delta)\,. \quad . \; . \; . \quad (3)$$

(It is approximate because both the damping factor δ and the coupling factor k are taken to be small, e.g. $\delta < 0.1$ and $k < 0.1$, which is usually the case.)

From the eq. (3) it is seen that the secondary circuit oscillates at its own angular frequency ω_0 (the cosine term) and that this oscillation is modulated with the much lower angular frequency $\frac{1}{2}k\omega_0$ (the sine term); this is also to be seen from the graphical representation in *fig. 10*. Further, the formula shows that the amplitude of the oscillation undergoes the natural damping (term with the e-power).

The term $\sin\frac{1}{2}k(\omega_0 t - \delta)$ corresponds to the energy flowing to and fro between the two circuits. In fig. 10 it is seen that the amplitude reaches a number of maxima. The ratio of two successive maxima depends upon the value of δ/k. The time τ between two maxima is given by $\frac{1}{2}k\omega_0\tau = \pi$, and hence:

$$\tau = \frac{2\pi}{\omega_0 k}\,.$$

The ratio a of a maximum in the envelope drawn in dotted lines (fig. 10) to the preceding maximum is therefore:

$$a = e^{-\frac{r}{2L}\cdot\frac{2\pi}{\omega_0 k}} = e^{-\frac{\pi\delta}{k}}\,. \quad . \; . \; . \; . \; . \; . \; . \quad (4)$$

Whereas from the shape of the frequency response curve of a band-pass filter it can immediately be seen whether the coupling is looser or tighter than the critical coupling (one peak or two peaks in the curve), the unit function response curve (fig. 10) does not provide any such simple criterion, since when passing the critical coupling this curve does not noticeably change in shape.

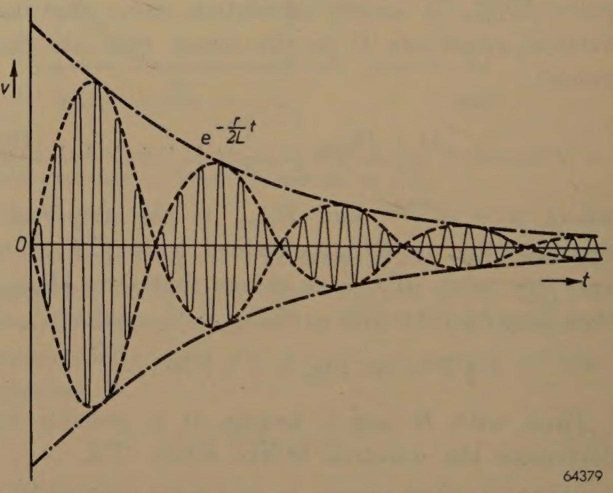

Fig. 10. Variation of the secondary voltage v of the band-pass filter in fig. 9 when a unit function current is applied to the primary.

This, however, does not prevent conclusions about the coupling being drawn from the oscillogram of the unit function response curve. According to eq. (4) the ratio a increases uniformly as the coupling becomes tighter and at the critical coupling assumes the value $e^{-\pi} \approx 1/_{23}$, that is to say, the second and following maxima are then already so small as to be almost indistinguishable in the oscillogram. If, however, the oscillogram shows marked and different maxima, as in fig. 10, then it is to be concluded that the coupling is very much tighter than the critical coupling. By measuring the ratio a from the oscillogram and applying formula (4) one can then determine with reasonable accuracy the ratio δ/k.

Low-pass filter

As a last example, we shall take the case of a low-pass filter, assuming that the output is terminated by a resistance R and that the input is open-circuited (*fig. 11*). The oscillograms obtained of the output voltage when the terminating resistance is too high, correct or too low are of the types represented in figures 11*a*, *b* and *c* respectively. If a resistance is applied also across the input of the filter and this resistance is approximately equal to the characteristic impedance of the filter, similar figures will be obtained but the oscillations are then much weaker. The most favourable value of the resistance is that with which the oscillations are as weak as possible, and with the aid of the oscilloscope this value can easily be determined experimentally.

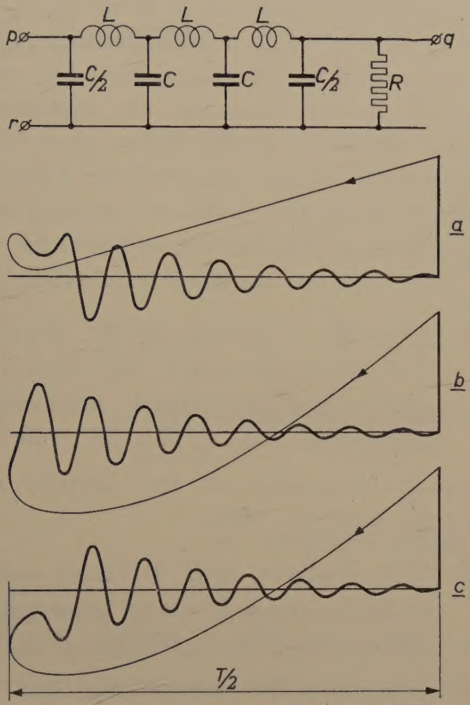

Fig. 11. *Top:* a low-pass filter terminated by a resistor R. When a unit function of current is applied to the input the output voltage varies according to a, b or c, depending upon whether R is too high, of the correct value, or too low respectively.

In the same way it is possible to investigate the effect of unequal filter sections, for which otherwise complicated calculations would be required.

Although these investigations by means of an oscilloscope and the unit function response method do not yield any very accurate quantitative results, this method has the advantage that it reveals very quickly which circuit elements in a particular case are of the most importance and which are less essential. Its application is particularly advantageous in small laboratories where the means available are restricted (both as regards equipment and in respect to personnel).

Summary. A simple experimental method is described for studying the response of an electrical network or that of an impedance to the sudden switching off of a constant direct current. The voltage produced by the switching off is observed as a function of time on the screen of a cathode-ray oscilloscope. The switching off is done periodically with the aid of a multivibrator, which supplies at the same time the time-base voltage for the oscilloscope. The voltage variation observed permits of a number of conclusions being drawn in regard to the network or the impedance, as several examples show. This method is particularly suitable in cases where an answer is required quickly and not too stringent quantitative requirements have to be met.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherland.

1932: G. H. Jonker and J. H. van Santen: Ferromagnetic compounds of manganese with perovskite structure (Physica **16**, 337-349, 1950, No. 3).

Various "manganites" of the general formula $La^{3+}Mn^{3+}O_3^{2-}$, $Me^{2+}Mn^{4+}O_3^{2-}$ (Me representing a bivalent metal) have been prepared in the form of polycrystalline products. Perovskite structures were found, i.a., for all mixed crystals $LaMnO_3$-$CaMnO_3$, for $LaMnO_3$-$SrMnO_3$ containing up to 70% $SrMnO_3$, and for $LaMnO_3$-$BaMnO_3$ containing less than 50% $BaMnO_3$. The mixed crystals with perovskite structure are ferromagnetic. Curves for the Curie temperature versus composition and saturation versus composition are given for $LaMnO_3$-$CaMnO_3$, $LaMnO_3$-$SrMnO_3$, and $LaMnO_3$-$BaMnO_3$. Both types of curves show maxima between 25% and 40% $Me^{2+}Mn^{4+}O_3^{2-}$; here all third electrons available contribute with their spins to the saturation magnetization. The ferromagnetic properties can be understood as the result of a strong positive Mn^{3+}-Mn^{4+} exchange interaction combined with a weak Mn^{3+}-Mn^{3+} interaction and a negative Mn^{4+}-Mn^{4+} interaction. The Mn^{3+}-Mn^{4+} interaction, presumably of the indirect exchange type, is thought to be the first clear example of positive exchange interaction in oxidic substances.

1933: R. Loosjes, H. J. Vink and C. G. J. Jansen: The potential distribution in pulsed oxide-coated cathodes and its consequences for the velocity distribution of the emitted electrons (J. Appl. Phys. **21**, 350-351, 1950, No. 4).

Previous investigations (see Nos **R 124** and **1910** of these abstracts) have shown that a potential difference of e.g. 200 V may exist across the oxide layer of an oxide-coated cathode at a current density of 10 A/cm² and that this potential drop is concentrated near the outer surface of the layer. In consequence of this the velocities of the emitted electrons show much greater spread than would be expected if the distribution were Maxwellian. The velocity spectrum of the electrically deflected electrons shows two (sometimes three) "lines" instead of being continuous, the spread amounting to 230 eV with $V_a = 860$ V, $I_a = 14$ A/cm², $T = 942$ °C.

R 137: G. Diemer and J. L. H. Jonker: On the time delay of secondary emission (Philips Res. Rep. **5**, 161-172, 1950, No. 3).

The frequency limit of a disc-seal dynatron proves to be 2400 Mc/s. From this an upper limit of 3×10^{-11} sec for the time lag of secondary emission is derived. Measurements on the admittance of an U.H.F. dynatron at 300 Mc/sec show that the time lag must be smaller than 10^{-11} sec. From transit time effects within the secondary-emitting material an upper limit for the time lag of the order of 10^{-14} to 10^{-15} sec may be estimated theoretically.

R 138: E. J. W. Verwey, P. W. Haaijman, F. C. Romeijn and G. W. van Oosterhout: Controlled-valency semiconductors (Philips Res. Rep. **5**, 173-187, 1950, No. 3).

Elaborate treatment of the subject already referred to under No. **1895**.

R 139: R. A. Hutner, E. S. Rittner and F. K. du Pré: Fermi levels in semiconductors (Philips Res. Rep. **5**, 188-204, 1950, No. 3).

General formulae for determining the Fermi level and the density of free charge carriers in semiconductors are derived. Special semiconductor models are considered in detail and a few applications are discussed.

R 140: J. M. L. Janssen: A cathode-ray oscillograph for periodic phenomena of high frequencies (Philips Res. Rep. **5**, 205-240, 1950, No. 3).

An oscillograph is described for periodic signals of high frequency, working according to a scanning principle. By mixing the high-frequency signal with phase-modulated pulses a faithful A.F. image is obtained. The bandwidth and the maximum scanning speed of the oscillograph are investigated as functions of the pulse frequency and pulse width. Further, the synchronization circuit, the mixer circuit and the pulse generator are described (c.f. Philips Techn. Review **12**, 52-59 and 73-82, 1950, No. 2/3).